

# UNIVERSITA' DEGLI STUDI DI PAVIA

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

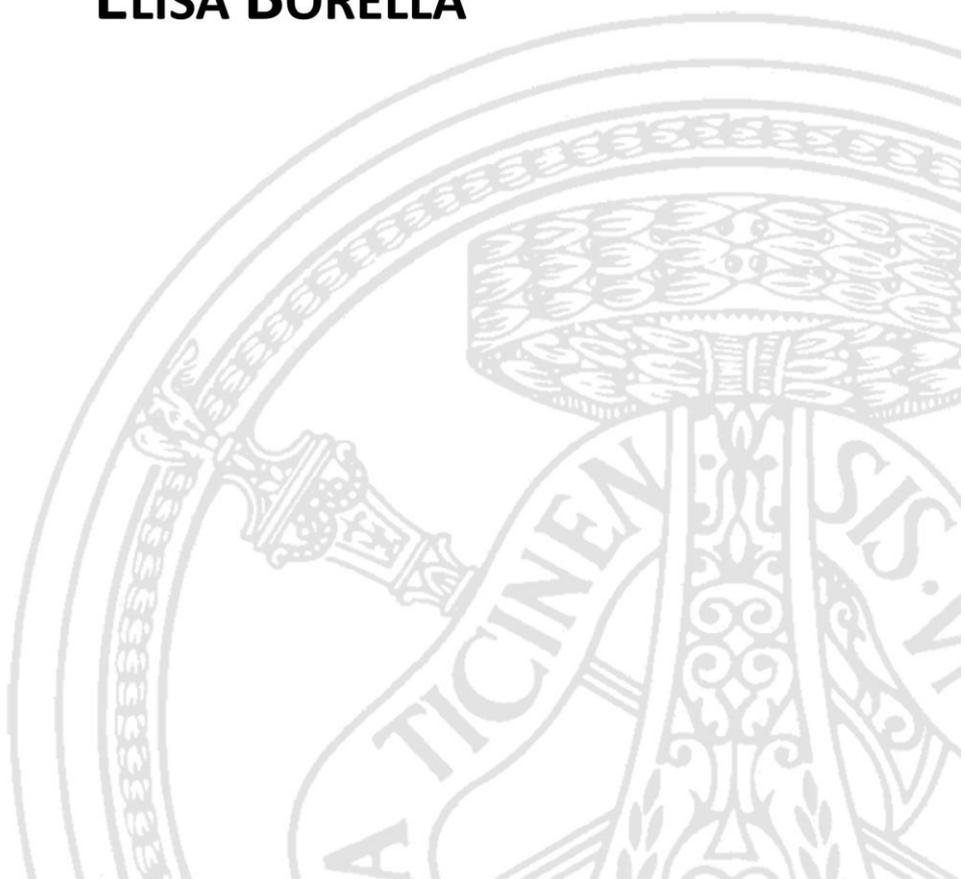
DOTTORATO DI RICERCA IN BIOINGEGNERIA E BIOINFORMATICA  
XXX CICLO

## **BAYESIAN METHODS AND TOOLS TO INTEGRATE PRIOR KNOWLEDGE IN THE DESIGN AND ANALYSIS OF PEDIATRIC CLINICAL TRIALS: A CASE STUDY ON BETA-THALASSEMIA, A RARE DISEASE**

PhD Thesis by  
**ELISA BORELLA**

**Advisor:**  
Prof. Paolo Magni

**PhD Program Chair:**  
Prof. Riccardo Bellazzi





*A Nicola, per essere sempre stato una fonte di supporto e incoraggiamento.*

*Ai miei genitori, per avermi insegnato con il loro esempio a lavorare  
duramente per raggiungere i miei obiettivi.*

---

# Acknowledgments

---

I would like to thank first my advisor, Prof. Paolo Magni, for his guidance and support during my Ph.D. I am grateful for all the opportunities he has afforded to me. He set such a great example for me with his commitment and hard work.

I would like to make a special reference to Prof. Oscar Della Pasqua for the opportunity to work in his Clinical Pharmacology & Therapeutics group, giving me the chance to meet an amazing group of people, work on a very interesting thematic, and expand my knowledge in the field.

I would like also to thank Prof. Cristiana Larizza, for her important contribution to part of this work: she has always had a solution to any WinBUGS's bug!

Thanks also to Dr. Italo Poggesi for all his insightful suggestions, for having always demonstrated a sincere interest in my work, and for always finding time for an advice.

Thanks to all the DDMoRe people who contributed to part of this work, especially Maciej Swat, Mike Smith, Richard Kaye, Gareth Smith, and Stuart Moodie.

Thanks to all the people of the Clinical Pharmacology & Therapeutics group for the amazing time spent in London. A special thank goes to Sean who helped me a lot to see things not only from an engineering perspective, to write better R code, and, last but not least, to say eighty-eight in Dutch.

Thanks to all the BMS lab people: Lorenzo, Nicola, Massimo, Ilaria, Letizia, Elisabetta, Silvia G., Roberta, Marilù, Enrica, Susanna, and, especially, to "le ragazze del corridoio", Elena and Silvia, who have shared with me few light but a lot of fun. I have found here an incredibly collaborative atmosphere and a wonderful group of people who I have enjoyed so much working with and learning from.

I would like to thank my friend Giulia: I know I can always count on you for anything! I would like to express my infinite gratitude to all my family: with their example, support, encouragement, and teachings, they made me what I am today and allow me to reach this achievement.

Finally, I would like to thank Nicola, for having always supported me, for all the little and big things you have done for me, for all the flights you took, for celebrating with me when even the smallest things went right and for always being there when things went less right. I was very fortunate for having you in my life.

---

## List of publications

---

### Publications on refereed journals

Terranova N, Rebuzzini P, Mazzini G, Borella E, Redi CA, Zuccotti M, Garagna S, Magni P. Mathematical modeling of growth and death dynamics of mouse embryonic stem cells irradiated with  $\gamma$ -rays. *Journal of Theoretical Biology*. 2014; 363:374-380.

Borella E, Poggesi I, Magni P. Prediction of the Effect of Renal Impairment on the Pharmacokinetics of New Drugs. *Clinical Pharmacokinetics*. 2017. doi: 10.1007/s40262-017-0574-9.

Carrara L, Lavezzi SM, Borella E, De Nicolao G, Magni P, Poggesi I. Current mathematical models for cancer drug discovery. *Expert Opinion on Drug Discovery*. 2017; 12(8): 785:799.

Lavezzi SM\*, Borella E\*, Carrara L, De Nicolao G, Magni P, Poggesi I. Mathematical modeling of efficacy and safety for anticancer drugs clinical development. *Expert Opinion on Drug Discovery*. 2017. doi: 10.1080/17460441.2018.1388369

Smith MK, Moodie SL, Bizzotto R, Blaudez E, Borella E, Carrara L, Chan P, Chenel M, Comets E, Gieschke R, Harling K, Harnisch L, Hartung N, Hooker AC, Karlsson MO, Kaye R, Kloft C, Kokash N, Lavielle M, Lestini G, Magni P, Mari A, Mentre F, Muselle C, Nordgren R, Nyberg H, Parra-Guillen ZP, Pasotti L, Rode-Kristensen N, Sardu ML, Smith G, Swat MJ, Terranova N, Yngamn G, Yvon F, Holford N. Model Description Language (MDL): A standard for modelling and simulation. *CPT Pharmacometrics Syst Pharmacol*. 2017. doi:10.1002/psp4.12222

### Conference abstracts

Borella E, Poggesi I, Magni P. Predictive assessments of pharmacokinetic alterations in subjects with renal disease. PAGE 24 (2015) Abstr 3442 [[www.pagemeeting.org/?abstract=3442](http://www.pagemeeting.org/?abstract=3442)]

Borella E, Carrara L, Lavezzi SM, Mezzalana E, Pasotti L, De Nicolao G, Magni P. Evaluation of software tools for Bayesian estimation on population models with count and continuous data. PAGE 24 (2015) Abstr 3452 [[www.pagemeeting.org/?abstract=3452](http://www.pagemeeting.org/?abstract=3452)]

Larizza C, Pasotti L, Mezzalana E, Borella E, Smith G, Magni P. Automatic translation of Bayesian pharmacometric models: the PharmML-to-WinBugs converter. PAGE 24 (2015) Abstr 3565 [[www.page-meeting.org/?abstract=3565](http://www.page-meeting.org/?abstract=3565)]

Borella E, Carrara L, Lavezzi SM, Massaiu I, Sauta E, Tosca EM, Vitali F, Zucca S, Pasotti L, De Nicolao G, Magni P. Methods and tools for multiscale modelling in Systems Pharmacology: a review. PAGE 25 (2016) Abstr 5793 [[www.page-meeting.org/?abstract=5793](http://www.page-meeting.org/?abstract=5793)]

Larizza C, Borella E, Pasotti L, Smith G, Kaye R, Magni P. Development and integration of the WinBUGS connector in the DDMoRe Interoperability Framework. PAGE 25 (2016) Abstr 5829 [[www.page-meeting.org/?abstract=5829](http://www.page-meeting.org/?abstract=5829)]

Tosca EM, Borella E, Terranova N, Rocchetti M, Magni P. Evaluation of a PK/PD DEB-based model for tumor-in-host growth kinetics under anticancer treatment. PAGE 25 (2016) Abstr 5875 [[www.page-meeting.org/?abstract=5875](http://www.page-meeting.org/?abstract=5875)]

Smith MK, Moodie SL, Bizzotto R, Blaudez E, Borella E, Carrara L, Chan P, Chenel M, Comets E, Gieschke R, Harling K, Harnisch L, Hartung N, Hooker A, Karlsson MO, Kaye R, Kloft C, Kokash N, Lavielle M, Magni P, Mari A, Mentre F, Muselle C, Nordgren R, Nyberg H, Parra-Guillen ZP, Pasotti L, Rode-Kristensen N, Smith G, Swat MJ, Terranova N, Yvon F, Holford N. Model Description Language (MDL) - a standard for model communication and interoperability. PAGE 25 (2016) Abstr 5899 [[www.page-meeting.org/?abstract=5899](http://www.page-meeting.org/?abstract=5899)]

Larizza C, Borella E, Pasotti L, Tartaglione P, Smith MK, Moodie SL, Magni P. Execution of complex Bayesian workflows with the DDMoRe Interoperability Framework: a case study in the diabetes area. PAGE 26 (2017) Abstr 7188 [[www.page-meeting.org/?abstract=7188](http://www.page-meeting.org/?abstract=7188)]

### **Conference oral presentation**

Borella E, Oosterholt S, Magni P, Della Pasqua O. Paediatric trial design optimization using prior knowledge in combination with modelling & simulations. Lewis Sheiner Student Session, PAGE 26 (2017) Abstr 7180 [[www.page-meeting.org/?abstract=7180](http://www.page-meeting.org/?abstract=7180)]

---

## Abstract (Italiano)

---

L'attività di ricerca descritta in questa tesi è stata condotta nell'ambito della modellistica matematica applicata allo sviluppo di farmaci. Il processo di registrazione e approvazione di nuovi farmaci viene tradizionalmente supportato da evidenze scientifiche sulla sua efficacia derivanti da studi clinici prospettici che coinvolgono centinaia, o addirittura migliaia, di pazienti. Questa tipologia di approccio al problema non è perseguibile quando si considerano sottopopolazioni di pazienti molto piccole, come nel caso di malattie rare che coinvolgono pazienti in età pediatrica. In questi casi, la difficoltà nel trovare un adeguato numero di pazienti arruolabili è spesso accompagnata da restrizioni sul numero di campioni biologici prelevabili per motivi di tipo metodologico ed etico. Al fine di garantire a queste popolazioni l'accesso a farmaci efficaci e sicuri è pertanto necessario integrare i dati raccolti durante gli studi con tutta la conoscenza a priori sul farmaco in studio e sulla malattia specifica, che possono derivare, per esempio, da precedenti studi sullo specifico farmaco pubblicati in letteratura e condotti in un'altra popolazione (per es. nella popolazione adulta) o su farmaci affini per meccanismo d'azione o indicazione terapeutica condotti sulla medesima popolazione di pazienti.

A supporto di questa necessità, in questa tesi, si è cercato di rispondere a tre importanti domande di interesse clinico e metodologico, utilizzando gli strumenti forniti dalla modellistica matematica e considerando un caso di studio reale, ovvero gli studi di farmacocinetica ed efficacia condotti nell'ambito del progetto europeo DEferiprone Evaluation in Paediatrics (DEEP) su farmaci chelanti del ferro in pazienti pediatriche affetti da emoglobinopatie rare.

Per prima cosa, è stato dimostrato in modo quantitativo quanto l'utilizzo di informazione a priori, attraverso l'applicazione di metodologie Bayesiane, permetta l'identificazione di modelli di popolazione in farmacocinetica anche quando i dati disponibili sono molto sparsi, come succede tipicamente in studi pediatrici che coinvolgono pazienti affetti da malattie rare.

In secondo luogo, sono stati investigati i potenziali benefici nell'utilizzare l'informazione a priori non solo a studio terminato, ma anche durante la pianificazione del disegno sperimentale dello studio clinico affinché i dati raccolti siano il più possibili informativi, sempre però considerando i vincoli pratici presenti quando si considerano queste popolazioni.

Inoltre, un'analisi di rischio-beneficio, confrontando diversi disegni sperimentali ottimizzati e non, ha permesso di quantificare i rischi connessi all'estrapolare conclusioni su efficacia e sicurezza di un farmaco quando si hanno pochi dati a disposizione e per giunta raccolti seguendo disegni di

campionamento non ottimizzati, come è tipico nel caso di studi pediatrici di farmaci per malattie rare.

Infine, a partire dalle conoscenze che si hanno sui meccanismi d'azione dei farmaci-chelanti del ferro, sulla regolazione dell'omeostasi del ferro e sui meccanismi che stanno alla base della patologia da accumulo di ferro, un modello di farmacocinetica-farmacodinamica è stato sviluppato e successivamente identificato a partire da soli dati disponibili dalla letteratura provenienti da studi su pazienti sia adulti che pediatrici affetti da emoglobinopatie rare. È stata poi investigata la possibilità di utilizzare questo modello per la predizione a lungo termine della risposta clinica di pazienti pediatrici, per permettere così un aggiustamento il più possibile tempestivo della loro terapia.

Per un utilizzo più intensivo della conoscenza a priori e dell'informazione storica occorre che tutta la comunità di modellisti sia facilitata nel condividere le proprie conoscenze e riutilizzare quelle altrui attraverso l'utilizzo di standard per la rappresentazione di dati, modelli, processi, etc.

A questo fine, il progetto europeo Drug Disease Model Resources (DDMoRe) ha recentemente rilasciato un'interoperability framework, ovvero una piattaforma modulare che integra diversi tool in un unico ambiente di programmazione, insieme ad un linguaggio standard di rappresentazione dei modelli che permette di eseguire con lo stesso codice di partenza task diversi (per es. di stima o simulazione) in tool diversi (per es. NONMEM, Monolix, SimulX, PopED).

In questa tesi, viene quindi anche presentata l'integrazione di un popolare tool per stima Bayesiana, WinBUGS, all'interno della DDMoRe interoperability framework, ed il suo utilizzo per la codifica di workflow complessi che comprendono l'esecuzione in diversi tool di task di stima e simulazione su due modelli di popolazione utilizzati nell'area di ricerca del diabete.

---

## Abstract (English)

---

The registration and approval of new compounds have traditionally been based on evidence arising from large prospective clinical trials. Such an approach is often not possible or unsuitable for small and very small populations as in case of rare diseases affecting children, where a limited number of patients is available and sparse samples are usually collected due to ethical and practical constraints. To ensure that there are adequate data to support the safe and effective use of drugs in these populations, a more efficient use of the evidence available from historical data is needed.

To this aim, in this thesis, three different important clinical questions have been addressed through the adoption of a model-based framework.

First, we have quantitatively demonstrated to what extent the use of historical data as prior knowledge, in conjunction with Bayesian statistical methods, can support the analysis of sparse data from new pediatric studies in a non-linear mixed effect modeling context.

Second, we have investigated to what extent integrating existing knowledge can support the optimization of study design (e.g. sampling times, sample size, trial duration) of future clinical trials. Besides, the possible risks in making inferences on drug safety and efficacy from trials characterized by very sparse and non-optimized designs have been quantitatively investigated.

Third, the use of a pharmacokinetic-pharmacodynamic model, developed from existing knowledge on the drug and the disease and subsequently identified on historical data, was evaluated for the prediction of patients long-term clinical response and, therefore, the adjustment of their ongoing therapy.

All these questions have been addressed using real-life examples of pediatric clinical trials, especially two inter-related trials on the treatment of transfusion-dependent hemoglobinopathies, in particular, the pharmacokinetic and the efficacy trials within the DEferiprone Evaluation in Paediatrics (DEEP) project. However, the proposed framework can be extrapolated to a broader range of diseases and conditions.

The basis for promoting the exploitation of prior knowledge and historical information is the availability of standards, tools, and processes common to all the modelling community that simplify knowledge sharing and reuse.

The Drug Disease Model Resources (DDMoRe) project has delivered an interoperability framework, integrating different currently-used tools, and exchange standards, which allow the user to write his model once and then to use it in all the different tools embedded in the framework, carrying out

different pharmacometrics tasks (e.g., estimation, simulation, design optimization).

To this aim, in this thesis, the integration of a currently-used tool for Bayesian estimation, WinBUGS, in the open-source DDMoRe interoperability framework, and its use into seamless standardized but flexible workflows, inspired by real-life modeling & simulations applications in the diabetes area, are presented.

---

# Contents

---

<b>Acknowledgments</b> .....	<b>IV</b>
<b>List of publications</b> .....	<b>V</b>
<b>Abstract (Italiano)</b> .....	<b>I</b>
<b>Abstract (English)</b> .....	<b>III</b>
<b>Contents</b> .....	<b>5</b>
<b>Chapter 1</b> .....	<b>7</b>
1.1. <i>Background</i> .....	7
1.2. <i>Thesis overview</i> .....	10
<b>Chapter 2</b> .....	<b>12</b>
2.1. <i>Introduction</i> .....	12
2.2. <i>Population PK modeling of deferasirox</i> .....	14
2.2.1. <i>Clinical PK data</i> .....	14
2.2.2. <i>Model building</i> .....	19
2.2.3. <i>Model evaluation and diagnostic criteria</i> .....	19
2.3. <i>DEEP-2 PK sub-study simulation: the original protocol</i> .....	20
2.3.1. <i>Simulation of a virtual pediatric population</i> .....	20
2.3.2. <i>Simulation of deferasirox PK data</i> .....	20
2.3.3. <i>Analysis of sparse PK data</i> .....	21
2.4. <i>DEEP-2 PK sub-study simulation: the optimized protocol</i> .....	22
2.4.1. <i>Optimal sampling</i> .....	23
2.4.2. <i>Simulation of PK data</i> .....	24
2.5. <i>Comparing original and optimized protocols</i> .....	24
2.6. <i>Results</i> .....	25
2.6.1. <i>Population PK modeling</i> .....	25
2.6.2. <i>Evaluation of the advantages of Bayesian estimation methods in analyzing sparse PK data</i> .....	26
2.6.3. <i>Evaluation of the impact of optimized PK designs in increasing precision for pediatric exposure extrapolation</i> .....	28
2.7. <i>Discussion</i> .....	32
<b>Supplementary material to Chapter 2</b> .....	<b>XXXV</b>
<b>Chapter 3</b> .....	<b>47</b>
3.1. <i>Introduction</i> .....	47
3.2. <i>Development of a PK-PD model for iron overload and chelation therapy</i> ....	51
3.2.1. <i>Clinical efficacy data</i> .....	51
3.2.2. <i>PK modelling</i> .....	52
3.2.3. <i>PK-PD modeling</i> .....	53

---

3.2.4. Results.....	54
3.3. <i>A model-based optimization of chelation therapy</i> .....	58
3.3.1. Simulation of a virtual pediatric population .....	59
3.3.2. Population PK-PD model.....	59
3.3.3. Original dosing adjustment strategy.....	60
3.3.4. Model-based dose adaptation strategy relying on long-term predictions.....	61
3.3.5. Results.....	63
3.4. <i>Discussion</i> .....	64
<b>Supplementary material to Chapter 3</b> .....	<b>LXVI</b>
<i>List of assumptions for dataset creation</i> .....	<i>LXVI</i>
<i>Additional figures and tables</i> .....	<i>LXX</i>
<b>Chapter 4</b> .....	<b>72</b>
4.1. <i>Model-based methodology to reduce trial duration</i> .....	73
4.1.1. Simulation of a virtual pediatric population .....	73
4.1.2. DEEP-2 non-inferiority study simulation: shorter trial durations.....	73
4.1.3. Comparing original and reduced trial durations.....	75
4.1.4. Results and discussion.....	76
4.2. <i>Model-based methodology to reduce sample size</i> .....	78
4.2.1. Non-inferiority assessment method based on proportions of success (method 1).....	79
4.2.2. Power calculation (method 1).....	80
4.2.3. Non-inferiority assessment method based on repeated measurement (method 2).....	81
4.2.4. Power calculation (method 2).....	82
4.2.5. Non-inferiority assessment method based on model-based approach (method 3).....	83
4.2.6. Results and discussion.....	84
<b>Chapter 5</b> .....	<b>88</b>
<b>Appendix A</b> .....	<b>91</b>
A.1. <i>Introduction</i> .....	91
A.2. <i>Methods</i> .....	95
A.2.1. Software.....	95
A.2.2. Implemented example workflow overview .....	95
A.2.3. Mathematical models.....	97
A.3. <i>Results</i> .....	100
A.3.1. Workflow results using a MLE approach.....	101
A.3.2. Workflow results using a Bayesian approach.....	101
A.3.3. Workflow results using a mixed approach .....	102
A.3.4. Supported features of the WinBUGS plugin .....	CVII
A.4. <i>Discussion</i> .....	CVIII
<b>Supplementary material to Appendix A</b> .....	<b>CX</b>
<b>Supplementary methods to Appendix A</b> .....	<b>CXIV</b>
<i>Software implementation: WinBUGS plugin and connector</i> .....	<i>CXIV</i>
<i>Description and assumptions for the automatically generated BUGS model</i> .....	<i>CXVIII</i>
<i>Supported operators</i> .....	<i>CXXVII</i>
<i>Supported probability distributions</i> .....	<i>CXXIX</i>
<b>List of Abbreviations</b> .....	<b>132</b>
<b>References</b> .....	<b>136</b>

---

# Chapter 1

---

## General Introduction and Thesis Overview

### 1.1. Background

Drug development is a challenging and resource-intensive field. An increasing difficulty in finding new molecular entities (NME) (1,2), together with a dramatic increase in the total costs to bring an NME to market (2,3), have been identified as the causes of the decline in pharmaceutical productivity (4). To effectively face these critical challenges, the pharmaceutical industry has been keen to stress the importance of augmenting the whole drug discovery process with mathematical modeling & simulation (M&S) (5). Despite most industries (e.g., automotive, aerospace, electronics) have routinely applied computer-aided M&S in their R&D processes since decades, the adoption of these techniques in pharmaceutical R&D has been slower since biological systems have been perceived as too complicated to be fully characterized with mathematical equations (4). However, despite a relatively late adopter, the picture changed in the last decade, and the term model-informed drug discovery & development (MID3) was popularized by industries as a solution to the decline in productivity (6). The use of M&S as a potentially valuable tool to improve drug development has been also encouraged by the European Medicines Agency (EMA) (7) and by the Food and Drug Administration (FDA) (8–10). To confirm this, FDA identified MID3 as a strategic component of the “Critical Path Initiative” established in 2004.

An essential part of MID3 is the discipline of pharmacometrics, which includes the development, using data collected in clinical trials, of mathematical models to characterize absorption, distribution, metabolism,

---

and excretion (ADME) properties of compounds (i.e. pharmacokinetics, PK), and the intrinsic drug activity or pharmacodynamics (PD), i.e. what the drug does to the body (5). Both elements are combined in PK-PD models.

Since humans differ from one another, the population approach was introduced in pharmacometric modeling. Non-linear mixed effect (NLME) modeling is one of the approaches used to account for variability not only at a data point level (i.e. unexplained or residual variability) but a parameter level, allowing parameters to vary among different individuals (i.e. inter-individual variability, IIV) (11). A NLME model is therefore characterized by a set of typical population model parameters, their variance in the population, and the variance of the residual unexplained variability. The general structure of a NLME model is as follows:

$$y_{ij} = f(X_{ij}, \varphi_i) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2) \quad (1.1)$$

where  $y_{ij}$  is the  $j$ -th observation for the  $i$ -th individual,  $f(\dots)$  is a non-linear function of a vector of individual parameters  $\varphi_i$  and a vector of independent variables  $X_{ij}$  (e.g., time, demographic covariates), and  $\varepsilon_{ij}$  is the difference between the individual prediction and the observation, i.e. the residual variability. In Eq. 1.1 an additive model has been used for the residual variability, but also a proportional model or a combination of the two is often adopted. The individual parameter  $\varphi_i$  for the  $i$ -th individual can be described by the expression:

$$\varphi_i = h(\theta, X_i) \cdot e_i^{\eta_i}, \quad \eta_i \sim N(0, \omega^2) \quad (1.2)$$

Where  $h(\dots)$  is a function of  $\theta$ , which is the typical value (or fixed effect) for the parameter  $\varphi$  in the population under study, and  $X_i$  is a set of individual covariates, while  $\eta_i$  is the random effect, that is the difference between the typical value and the individual parameter value. In Eq. 1.2 the random effect is assumed to be log-normally distributed with respect to the fixed effect, but also other models that relate the random to the fixed effects are possible.

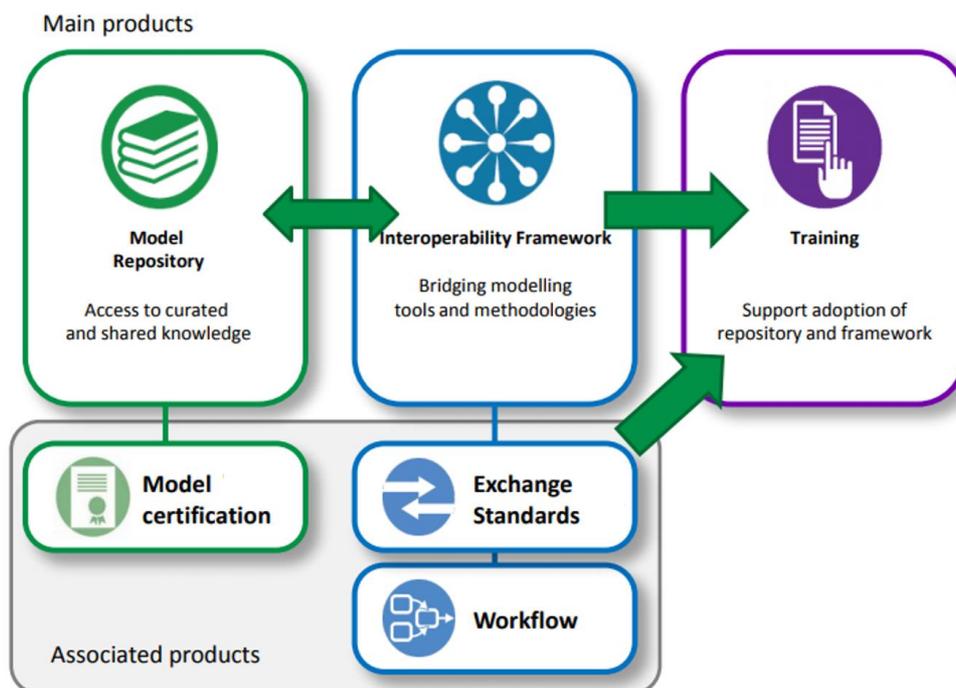
Evidence of PK-PD relationships for a new compound requires data arising from well-designed and relatively large clinical trials. Lack of attention to such requirements usually yields PK and PD parameters estimates which are biased or show poor precision. However, this prerequisite is often not possible in case of rare diseases, where the patient population is small and the clinical experience dispersed among a small number of clinical referral centers. The difficulties become even more acute for pediatric orphan drugs (i.e. drugs to treat rare diseases), where patient recruitment, retention, and management present more challenges because of patients younger demographic, e.g., family dynamics, ethical and practical constraints. A careful trade-off between reducing risk and discomfort, and obtaining informative data is therefore required (12).

In these cases, effective methods for data analysis and study design optimization should be used, including Bayesian estimation methods and optimal design techniques. Both allow the incorporation of prior knowledge, deriving from historical data from similar trials in adults or pediatric populations. Optimal design is a technique where a previous or assumed PK-PD model is used to define which is the most informative design so that data collected will give precise estimates of model parameters, thus a proper characterization of PK-PD relationships. Optimal designs methods are based on the optimization of a function of the Fisher information matrix (FIM) (e.g. the determinant for the D-optimality criterion). Different design factors can be optimized, e.g., number of patients, number of samples, sampling times, and study duration. Bayesian approaches can be incorporated to take into account uncertainty in parameter values in the optimization, by employing the ED-optimality design criteria; priors can be assumed on each parameter, deriving them from previous studies on the same drug or another drug in the same class (13).

To enable hypotheses testing and to probe potential outcomes (e.g., what is the impact of an optimized sampling schedule on the uncertainty of model parameters? What is the effect of collecting more samples?), M&S, and, in particular, clinical trial simulations (CTS), are invaluable tools in decision-making. CTS combine different types of models to simulate the outcome of a clinical trial, generating in output a dataset similar to the actual data coming from a real trial. Many replicates of trials can be simulated to allow statistical analysis of simulation results, with the final aim of testing specific questions, ‘improving the quality, efficiency and cost-effectiveness of decision making’ (6). To do that, a PK-PD model describing the disease progression and the drug effects on the disease, a model for the covariate distribution in the study population, and a model for the study design (e.g., dosing regimen, treatment arms) are necessary (13). Therefore, using CTS, a variety of different scenarios can be simulated and compared, and the impact of different design elements evaluated.

To reach the goal of rapid access to safe and effective pediatric orphan medicines is therefore of crucial importance the integration of all relevant prior knowledge to support the design/analysis of studies/experiments. This could benefit from knowledge- and model-sharing. Sharing knowledge through sharing computer code is difficult since various software tools for estimation tasks, optimal design, and CTS can be employed, each of them having its model formulation/language, capabilities, methods, and algorithms. This lack of common standards hampers knowledge transfer, impacts reproducibility, and forces the users to manually translate the code into the one specific of the tool being used to be able to integrate it into existing workflows. This perspective introduces the activities of the Drug Disease Model Resources (DDMoRe) consortium ([www.ddmore.eu](http://www.ddmore.eu)), founded in 2011 under the Innovative Medicines Initiative (IMI) and ended in 2016, whose strategy has standards as its core. In fact, integral parts of

this project have been newly developed exchange standards for model definition (Modeling Description Language, MDL), software interoperability (Pharmacometrics Markup Language, PharmML), and software output (Standard Output, SO); an open-source interoperability framework (IOF), integrating different currently-used tools (e.g., NONMEM, WinBUGS, Monolix, PopED, PFIM, R); a publicly available Model Repository (repository.ddmore.eu) where model codes can be stored, retrieved, and shared with the community (Figure 1.1).



**Figure 1.1:** DDMoRe main and associated products. Adapted from (14).

In conclusion, the application of MID3 concepts is of key importance for pediatric orphan drug discovery and development. In particular, an efficient design of pediatric trials and an efficient analysis of emerging data using approaches informed by prior knowledge (e.g., ED-optimal design, Bayesian estimation methods) should be encouraged. DDMoRe and its delivered products allowed the construction of a quantitative framework to strengthen data sharing and enhance integration, motivating the user community to increase capabilities and capacities in the delivery of MID3.

## 1.2. Thesis overview

Following the above considerations, the aim of this thesis is to investigate how prior knowledge integration can impact on analysis and design of pediatric clinical trials, especially in case of rare diseases. This investigation

uses beta-thalassemia as a paradigm for a rare disease, and the DEEP-2 efficacy study, including a PK sub-study, conducted by the DEferiprone Evaluation in Paediatrics (DEEP) consortium (15), as a real-life example. This thesis project has been carried out at the Bioinformatics, Mathematical Modelling and Synthetic Biology (BMS) laboratory of University of Pavia and, in part, at the Clinical Pharmacology & Therapeutics group of UCL School of Pharmacy (London, UK), in collaboration with Prof. Oscar Della Pasqua.

The dissertation is organized as follows. In **Chapter 2** CTS of the DEEP-2 PK sub-study were performed to evaluate to what extent the use of prior knowledge and Bayesian estimation methods can support the analysis of sparse samples collected in a very limited number of children. Besides, the use of prior knowledge in the optimization of the study design was investigated for the DEEP-2 PK sub-study and CTS were performed to evaluate the impact of optimized sampling, and to probe potential outcomes, whether beneficial or deleterious, of alternative designs.

In **Chapter 3** the use of prior knowledge in combination with a newly developed PK-PD model for iron overload is evaluated to predict long-term clinical response and optimize chelation therapy in pediatric patients affected by beta-thalassemia and other transfusion-dependent hemoglobinopathies.

In **Chapter 4** CTS of the DEEP-2 efficacy study were performed to assess the possibility of reducing the trial duration and sample size using a model-based approach, based on the PK-PD model for iron overload developed from historical data and described in Chapter 3. An overall conclusion is reported in **Chapter 5**.

Finally, in **Appendix A** the development of a new WinBUGS plugin and its integration in the DDMoRe IOF is presented. This plugin facilitates the encoding and execution of complex Bayesian workflows. The plugin was tested in an advanced real-world case study, involving two inter-related models publicly available on the DDMoRe model repository.

---

# Chapter 2

---

## Use of prior knowledge for the design and analysis of pediatric PK trials for small rare disease populations

### 2.1. Introduction

To define the appropriate dose rationale and dosing regimens to be used in pediatric trials, it is important to characterize drug PK. This characterization is full of challenges, especially in the context of pediatrics and rare diseases where the number of patients available is small and the number of specialized treatment centers and pediatric clinical research infrastructures very limited. In addition, practical and ethical constraints to clinical testing in children lead, for example, to the collection of sparse data (1-3 samples per subject) and to constraints on the sampling windows (16,17), limiting the use by clinicians of the results of these studies as dosing guidelines.

These limitations constrain physicians to extrapolate from the adult population and to adjust dosing regimens, by simply scaling the dose to a child's body weight (or body surface area, BSA) assuming a linear relationship between weight and drug exposure. However, it is well known that during the childhood all the processes involved in the PK of a drug, from the absorption to the elimination, nonlinearly change with the weight increase (18).

NLME modeling is one of the preferred tools to analyze data coming from studies with fewer samples per individual and small sample size, especially when historical data are integrated in a parametric way (19). In NLME modeling, in fact, the population is analyzed as a whole, considering the data originated from different patients and there is also the possibility to pool

historic data of previous PK studies with sparse data coming from novel studies characterized by few observations per patient.

Prior PK data can be helpful not only for data analysis but also to optimize data collection itself, allowing a more rational design of experimental protocols. ED-optimality sampling method allows defining the best sampling times able to provide enough information for the PK model parameter estimates, by minimization of their standard errors, starting from prior PK data (20,21).

Despite the availability of different methods to integrate prior knowledge (e.g. Bayesian methods) for the analysis of such sparse and unbalanced data (22), and for the optimization of the study design (23,24), the application in clinical research, and, particularly, in case of rare diseases, is limited. Some attempts have been made recently in (25) where a model-based approach was used to characterize the systemic exposure of deferiprone in patients aged < 6 years old. In the work of Petit et al., a pediatric PK study for an antimalarial drug was designed based on information about adults and a recommended design was optimized on the basis of an extrapolated model built from historical data (26).

If historical data are not integrated into the analysis of data coming from small pediatric studies, there is a high risk that no reliable conclusions can be drawn on the drug under investigation due to a lack of statistical power, as well as a high false-negative rate (27).

These considerations lead to the objectives of the investigation described in this Chapter, in which the DEEP-2 PK sub-study (15), aimed to characterize the PK profile of deferasirox in pediatric iron-overload patients affected by rare hemoglobinopathies, was used as case study to evaluate:

(i) to what extent prior knowledge (e.g. previous studies in adults), in conjunction with Bayesian estimation methods can improve parameter estimation from data sparsely collected in a limited number of subjects;

(ii) how prior knowledge, together with optimality concepts can be used to ensure a more efficient design of sampling schemes in children.

This work expands on previous efforts in developing suitable methodologies for the evaluation of rare diseases (25,28,29).

Recently, in fact, the DEEP consortium, an initiative under the auspices of the FP7 program, has been established to evaluate the use of deferasirox and deferiprone, two chelating agents, in pediatric patients affected by hemoglobinopathies (15).

Since the limited evidence was available on the use of deferasirox and deferiprone in the pediatric population, PK was investigated to characterize drug exposure and confirm the effect of demographic covariates on the disposition of the investigational products.

For deferiprone, such an evaluation was undertaken in a separate PK study (DEEP-1 PK study), which showed the impact of both sex and creatinine clearance on the exposure (85). DEEP-1 PK study details are reported in Table 2.S1 in *Supplementary material to Chapter 2*. Complete information

---

on the DEEP-1 study protocol can be found in <https://www.clinicaltrialsregister.eu/ctr-search/trial/2012-000658-67/IT#E> (Italy) and in <https://www.clinicaltrialsregister.eu/ctr-search/trial/2012-000658-67/3rd> (outside EU/EEA).

The DEEP-2 study is an efficacy and safety study to evaluate non-inferiority of deferiprone compared to deferasirox in pediatric patients from 1 month to 18 years old, affected by hereditary hemoglobinopathies and requiring chronic transfusion. Within the DEEP-2 study, a PK sub-study in a subgroup of patients treated with deferasirox was conducted (Table 2.S1 in *Supplementary material to Chapter 2*). Complete information on the DEEP-2 study protocol can be found in <https://www.clinicaltrialsregister.eu/ctr-search/trial/2012-000353-31/IT> (Italy), in <https://www.clinicaltrialsregister.eu/ctr-search/trial/2012-000353-31/3rd> (outside EU/EEA) and in <https://www.clinicaltrialsregister.eu/ctr-search/trial/2012-000353-31/GB> (UK).

In the DEEP-2 PK sub-study a maximum of one plasma sample was considered for each patient ( $n = 19$ ) due to feasibility criteria. Sampling was done at the end of treatment (i.e. at steady state) to allow further characterization of the relationship between exposure and ferritin response (Table 2.S2 in *Supplementary material to Chapter 2*). A sampling window of four hours after the last dose was considered, based on a preliminary analysis of deferiprone. Each subject was randomly assigned to one of ten different sampling times: 15 minutes pre-dose, and 15, 30, 45, 60, 75, 90, 105, 120, 240 minutes post-dose. A deviation of  $\pm 10$  minutes was allowed around each sampling time.

Given important practical limitations associated with blood sampling in this patient population, we explore opportunities for ensuring efficient data collection and evaluate prospectively alternative methodologies to deal with data sparseness, uncertainty and (poor) precision in parameter estimation.

The possibility of estimating, in the pediatric population, individual primary and secondary PK parameters with sufficient precision, using only one sample per child represents, therefore, a key aspect of the methodology proposed in the subsequent paragraphs.

To this purpose, first, the PK profiles of deferasirox in a virtual population of pediatric patients were simulated using a PK model developed from historical data, appropriately scaled; subsequently, the primary and secondary PK parameters estimated from a subset of simulated data were compared to the ones used for simulation.

## **2.2. Population PK modeling of deferasirox**

### **2.2.1. Clinical PK data**

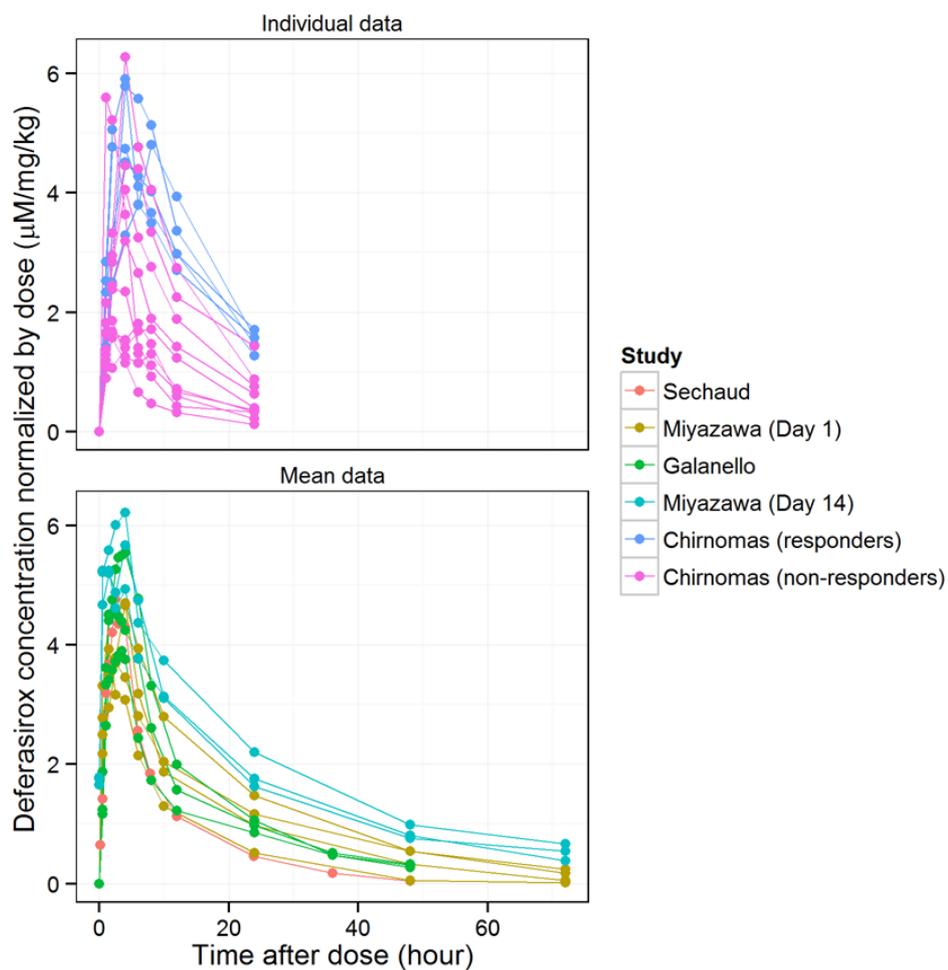
Data from five published PK studies with deferasirox were extracted from the literature (30–34) (Figure 2.1). An overview of these studies used for the model-based meta-analysis is shown in Table 2.1, including details about the study protocols and patient populations.

Studies were included in the analysis if the following criteria were met:

- single or multiple oral dose(s) of deferasirox were administered within the study period;
- the investigated population consisted of healthy adult individuals and adult or pediatric patients with transfusion-dependent hemoglobinopathies of both genders and of different races, and the reported data must include mean or individual time-concentration data;
- the year of publication of the selected studies ranged from 2006 to 2010.

To ensure appropriate data aggregation, data from the different sources had to be normalized based on the assumption of linear PK of deferasirox in the range of observed doses.

First, all doses were converted into  $\mu\text{mol}$ , using a molecular weight of deferasirox of 373.362 g/mol (35). While accurate information about the dose (amount) is critical for the characterization of PK parameters, assumptions had to be made when individual patient-level data were not reported. Since the dose is often expressed in mg/kg, mean body weight was used -when reported- to calculate the actual dose in mg. When body weight was missing, assumptions were made about the mean or individual body weight based on reported demographics such as age, sex, and race. All assumptions are listed in Table 2.1.



**Figure 2.1:** Concentration vs. time profiles of deferasirox normalized by dose. Individual profiles from Chirnomas et al. (30) (responders and non-responders) in the upper panel. Mean data (of both healthy and patients with different types of hemoglobinopathies) from the other publications (31,32,34) in the lower panel.

Reference	Daily dose	Sampling times	N° of patients	Age (yrs)	Weight (kg)	Sex (male:female)	Race	Disease	Exclusions	Additional comments
(31)	5 mg/kg 10 mg/kg 20 mg/kg 30 mg/kg	Samples on Day 1 and Day 14	6 7 6 7	71.5 68.0 66.0 75.0	Not reported. Mean weight for Asian regions reported in (36) was assumed.	1:5 3:4 1:5 3:4	Japanese	3 MDS, 3 AA, 0 other 5 MDS, 1 AA, 1 other 4 MDS, 1 AA, 1 other 4 MDS, 1 AA, 2 other	Data at Day 14 relative to the 30 mg/kg study arm excluded	Data excluded because of high noise in the data
(32)	375 mg	Samples on Day 1 at pre-dose (0), 15, 30, 45 min, and at 1, 1.5, 2, 2.5, 3, 3.5, 4, 6, 8, 12, 24, 36, and 48 h post-dose	17	30.5	79.05	17:0	22.2% Caucasian, 16.7% black, 61.1% other race.	Healthy volunteers	-	-
(33)	1000 mg (~20 mg/kg)	Samples on Day 7 at pre-dose (0), 0.5, 1, 1.5, 2, 3, 4, 6, 8, 12, 24, 48, 72, 96, 120, 144, 168 h post-dose	5	20-38	50-81	3:2	Not reported	Thalassemia	Data excluded	Data excluded because discrepancies between the dose reported in mg/kg, the mean body weight of the study population and the actual dose administered in mg

Reference	Daily dose	Sampling times	N° of patients	Age (yrs)	Weight (kg)	Sex (male:female)	Race	Disease	Exclusions	Additional comments
(34)	20 mg/kg	Samples on Day1 at pre-dose (0), 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 6, 8, 12, 24, 36, 48 h post-dose	28	18-45	A mean weight of 70 kg was assumed.	28:0	Not reported	Healthy volunteers	Data relative to arm C excluded	Difficulties in extracting data since measurements are overlapping with the other treated arms
(30)	34.7 mg/kg	Samples on Day1 at pre-dose (0), 1, 2, 4, 6, 12, 24 h post-dose	15	9-38 3-36	Patients < 18 years: calculated from individual age, sex, and race using the growth charts in (37). Patients > 18 years old: a mean weight of 70 kg was assumed.	3:2 7:3	1 Asian, 1 black, 3 white 4 Asian, 1 black, 5 white	1 SCD, 4 thalassemia 1 SCD, 9 thalassemia	-	-

**Table 2.1:** Description of clinical pharmacokinetic (PK) studies. MDS: myelodysplastic syndromes; AA: aplastic anemia; SCD: sickle-cell disease.

### 2.2.2. Model building

During model building one and two-compartment structural models with first-order absorption were tested. IIV was introduced on each PK parameter using an exponential structure in a stepwise manner.

Since only oral data were available, the bioavailability was fixed to 0.7 as reported in the work of Sechaud et al. (32).

Proportional, additive and combined error models were evaluated.

Given the historical understanding and the limited information available in literature, allometric scaling was inferred as the main covariate factor. Consequently, an allometric covariate structural model was applied to clearances (with a fixed exponent of 0.75) and volumes (with a fixed exponent of 1) and a reference adult weight of 70 kg. Mean body weight was considered when only PK mean profiles were available, while individual body weight was derived for each patient from individual age, sex, and race (Table 2.1) when individual PK profiles were available.

No maturation has been considered on clearances since all the patients participating to the PK sub-study were at least one year old, and deferasirox is mainly metabolized via glucuronidation by the UDP-glucuronosyltransferase (UGT) 1A1 to the acyl glucuronide metabolite, whose activity increases after birth and reaches adult levels by six months of life (38–40). Oxidative metabolism by cytochromes p450 (CYP) 1A2 and 2D6, which are known to reach adult values at one year of life (41,42), affects only the 10% of the dose (33,43).

The following estimation procedures were used to identify model parameters: population parameters (named in the following as THETAs) were estimated using only mean data (Table 2.1). IIV parameters (named in the following as OMEGAs) were estimated using only individual data of responders and non-responders from the study reported in Chirnomas et al. (30) (Table 2.1), fixing THETAs to the previous estimates.

### 2.2.3. Model evaluation and diagnostic criteria

Selection of the best model was based on objective function values, completion of estimation and covariance steps, precision of the parameter and error estimates, number of significant digits, correlation between parameters, and absence of zero gradients.

Besides, a visual inspection of goodness-of-fit (GOF) plots, such as population and individual predicted vs. observed concentrations, and conditional weighted residual (CWRES) vs. observed concentrations (or time), was carried out.

The final model was further evaluated by conducting a bootstrap, visual predictive checks (VPC) and normalized predictive distribution errors (NPDE) for both mean and individual data. The VPC for mean data was

---

obtained calculating the median, the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the means of the simulated data.

Model building and evaluation were performed using NONMEM v.7.3 (Icon Development Solution, USA). Conditional estimation with interaction was used as estimation method.

R v.3.0.3 was used for GOF plots, NPDE, and VPC.

## **2.3. DEEP-2 PK sub-study simulation: the original protocol**

The first objective of this work was to support and improve parameter estimation in studies where conventional options for study optimization may not be feasible (e.g., more patients and/or more samples/patient) as in the DEEP-2 PK sub-study. We wanted to point out the possible limitations of the original design and clarify expectations on the reliability of parameters estimates.

### **2.3.1. Simulation of a virtual pediatric population**

Since the PK model for deferasirox includes a covariate effect (weight) on clearances and volumes, a virtual patient population with a representative covariate distribution was simulated.

A large number (15,000) of hypothetical patients were considered, aged from 1 to 18 years old, with a 1:1 sex ratio. Corresponding weights were simulated using an appropriate demographic model (44), which includes correlations between postmenstrual age (PMA), sex, and body weight. To simulate body weight, PMA was extracted from a uniform distribution between 1 year and 18 years old (plus a gestational time of 40 weeks).

### **2.3.2. Simulation of deferasirox PK data**

Using the PK model built on historical data and scaled to pediatric patients (Table 2.2), deferasirox plasma concentration-time profiles at steady state were generated assuming a maintenance dose of 20 mg/kg.

Different scenarios were simulated with the goal of identifying the most suitable method for the analysis of clinical trial data:

- Scenario 1) the PK of deferasirox is assumed to be accurately described by the proposed PK model, therefore deferasirox concentration-time profiles of the 15,000 virtual pediatric subjects were simulated starting from the same population parameters estimated from literature data (Table 2.2).
- Scenarios where the actual PK of deferasirox is assumed to vary significantly from the profiles predicted by the proposed PK model:

- Scenario 2) CL was significantly lower than predicted (i.e. 50% of the predicted value);
- Scenario 3) Both CL and V2 were significantly lower than predicted (i.e. 50% of the predicted value);
- Scenario 4) All disposition parameters (CL, V2, Q, and V3) were significantly lower than predicted (i.e. 50% of the predicted value);
- Scenario 5) CL and Q were calculated using an exponent of 0.85 (rather than 0.75) in the allometric equation;
- Scenario 6) CL and Q were calculated using an exponent of 2/3 (rather than 0.75) in the allometric equation.

Scenarios 2 to 4 accounted for the possibility that allometric scaling was not adequate to fully explain the differences between the adult and the pediatric population. Scenarios 5 and 6 accounted for the possibility that the allometric exponent chosen was not the right one. Six datasets, one for each scenario above illustrated, containing deferasirox levels for 15,000 virtual patients, were obtained by simulation. Then, from each of these simulated datasets, 19 subjects (i.e. the number of subjects enrolled in the original PK sub-study) were extracted randomly (with equal probabilities). Then, only one sample for each subject was extracted randomly (with equal probabilities) among the 10 sampling times defined in the original protocol (with a possible deviation of +/- 10 minutes).

Since in the pediatric population the body size-normalized renal function and liver's metabolic activity are decreased compared to adults, only scenarios where the population PK parameters are reduced compared to predicted will be considered hereafter. Scenarios where the population PK parameters in the pediatric population are higher than predicted were also tested but they will be not shown here.

The conditions mentioned above were identified among a range of possible perturbations, which were deemed sufficiently robust to assess the sensitivity of different methods for the estimation of PK parameters.

### **2.3.3. Analysis of sparse PK data**

From the simulated dataset the following PK model parameters were estimated: population parameters for CL, V2, V3,  $k_a$ , and Q, and IIV parameters for CL, V2, V3, and  $k_a$ . Conditional estimation with interaction was used as estimation method for each scenario.

Three approaches have been used: population analysis without priors, with highly-informative priors or weakly-informative priors.

Both highly- and weakly-informative priors were set via the \$PRIOR option in NONMEM. With \$PRIOR the prior information about the PK model parameters that was derived from the literature data is combined with the information coming from the sparse data of the new study, contributing to overcome difficulties in parameter estimation. This approach is close to

---

the Bayesian estimation approach, but it is not itself purely Bayesian. The \$PRIOR can be considered an extension of the classical frequentist approach, in which each parameter is always considered an unknown constant to be estimated, but it is supposed to have a certain probability density  $dR$ , so that  $O^P = -2 \cdot \log(dR)$  is a penalty term that is summed to the objective function  $O_S$  computed from sparse data. Upon minimizing the objective function  $O^P + O_S$ , for each value of the considered parameter, the less likely this value is, the greater is the “penalty” paid for considering this value as a possible point estimate (45).

In NONMEM, since THETAs are the population means of normally or lognormally distributed individual parameters, their priors have a normal density, either on a natural or log scale. Priors on OMEGAs have an inverse-Wishart density, since they are construed as variance-covariance of normally distributed individual parameters, either on a natural or log scale.

In the approach with highly-informative priors mentioned above, normal priors were used on THETAs with mean equal to the estimates reported in Table 2.2 and variance equal to  $SE^2$  ( $SE$ =standard error), considering only the estimation uncertainty. Inverse-Wishart priors were used on OMEGAs with mode equal to the estimates reported in Table 2.2 and degrees of freedom (DF) calculated as:

$$DF = 2 \cdot \left( \frac{OMEGA}{SE} \right)^2 + 1. \quad (2.1)$$

In the approach above with weakly-informative priors, higher variances of the normal priors on THETAs were chosen: the variance (expressed on the %CV scale) was set to 40% for CL and V2, and to 80% for Q, V3 and  $k_a$  (Figure 2.S1 in *Supplementary material to Chapter 2*). These values were chosen according to this empirical rule: %CV was set to 40% if  $5 \cdot RSE$  ( $RSE$ =relative standard error) of the parameter was below 40%, while was set to 80% if  $5 \cdot RSE$  was between 40% and 80%.

For all six scenarios and for each of the tested approaches (i.e. population analysis with no priors, highly-informative priors, and weakly-informative priors) the procedure described above was repeated until 100 successful runs were obtained. A run was considered successful if the minimization routine terminated successfully without any rounding errors. Covariance step was not considered at this point because has a high probability to fail with sparse data (more than 50% according to (45)); besides, the reliability of the estimated values was assessed comparing each estimated individual parameter with its actual value (i.e. the value used in simulation).

## 2.4. DEEP-2 PK sub-study simulation: the optimized protocol

### 2.4.1. Optimal sampling

A second objective of the present investigation was to identify best practice in clinical PK studies for rare diseases. Of interest is the need to demonstrate to what extent feasibility considerations can lead to biased estimates of the parameters of interest, and most importantly, how more informative a PK study can be if optimal design principles are applied. These principles remain overlooked and are undervalued by the clinical research community.

Here, we used ED-optimality principles given its relevance for situations where uncertainty is a critical component of the optimization procedure.

ED-optimization of the sampling times was performed with the PopED R package (development version) (20,21), using the line-search method. An uncertainty on the model parameters was taken into account assuming the following distributions: a zero-truncated normal distribution around each population parameter with mean equal to the estimated value in Table 2.2 and variance equal to  $SE^2$ , and an inverse-Wishart distribution around each IIV parameter with mode equal to the estimated value in Table 2.2 and DF calculated as in Eq. 2.1.

The covariate body weight was considered as a parameter (not to be estimated) that can assume different values among the 19 subjects according to a uniform distribution with median equal to the median of the weights calculated on the full dataset of 15,000 hypothetical patients (40 kg), and length equal to the difference between the 5<sup>th</sup> and 95<sup>th</sup> percentiles (30 kg).

The sampling schedule (i.e. when to take the samples) was set as the design parameter to be optimized, the number of subjects was fixed to the original number, 19, because the goal here was to evaluate the added value of optimizing the sampling protocol. Firstly, the number of samples per subject was fixed to one and the minimal and maximal sampling times were fixed to one hour pre-dose and four hours post-dose, respectively.

Secondly, to find the optimal number of sampling times per subject, several designs characterized by 2, 3, 4, 6, 8, 10 sampling times were compared in terms of precisions of model parameters. The scenario with the minimal number of different sampling times that was able to guarantee a sufficient precision of model parameters (i.e.  $CV% < 30%$ ) was chosen and used in the subsequent analysis. Group of subjects sharing the same sampling time will be defined hereafter with the term “design group”.

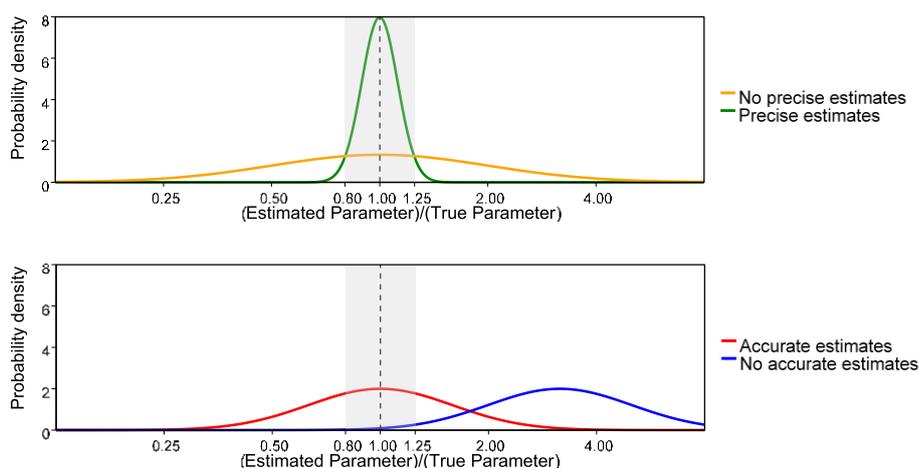
Many optimization runs have been performed; for each run, a different set of parameters is extracted by the ED-algorithm from their corresponding distributions and a different set of optimal sampling times is finally returned. In this way, a probability distribution of the optimal sampling time chosen for each design group was derived. To obtain smooth probability distributions profiles, 200 iterations of the optimization algorithm were necessary. The optimal sampling windows were then calculated as the narrowest intervals covering the 50% of each probability distribution.

## 2.4.2. Simulation of PK data

The simulation-estimation procedure described in paragraph 2.3.3 was performed again, always extracting 19 subjects, but in this case to each subject was assigned one sample as well as two or more samples chosen within the optimal sampling windows. Given restrictions regarding the maximum blood volume per day up to four samples per patient were considered.

## 2.5. Comparing original and optimized protocols

The scenarios presented were compared in terms of probability of successful convergence of the NLME algorithm, calculated as the ratio between 100 and the number of runs necessary to obtain 100 successful runs. The precision of the estimates was assessed via the probability distributions of the ratios of each estimated individual parameter to its ‘true’ value (the one used for simulating data), calculating for each of them the proportion of the density area between 1/1.25 and 1.25. A variation of 25% from the nominal value is usually considered a plausible variation from a biological perspective (46); for this reason (1/1.25; 1.25) was used as a range, in which the estimates of the exposure-related parameters can be considered reliable. The probability density of the ratios of estimated individual parameter values to their true values is expected to be centered in one (see red curve in Figure 2.2) if individual estimates are accurate. A deviation from one indicates the presence of bias (see blue curve in Figure 2.2). Besides, more the probability density curve is narrowed around one, more the estimated parameter values can be considered precise (see green curve compared to orange curve in Figure 2.2).



**Figure 2.2:** Probability distributions of ratios of the estimated parameter to the original parameter (used in simulation). The upper panel shows the probability density curves of precise (green) vs. less-precise (orange)

estimates. The lower panel shows probability density curves of accurate (red) vs. less-accurate (blue) estimates.

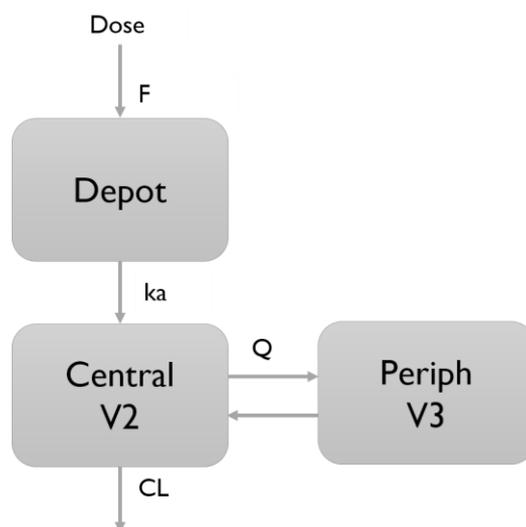
## 2.6. Results

### 2.6.1. Population PK modeling

The PK of deferasirox was best described by a linear two-compartmental model with first-order absorption and elimination (Figure 2.3). IIV was estimated for clearance (CL), volume of distribution of the central compartment (V2), absorption rate constant ( $k_a$ ) and volume of distribution of the peripheral compartment (V3). Residual variability was characterized by a proportional error model.

Final parameter estimates (with bootstrap median and 90% confidence interval) are reported in Table 2.2.

Diagnostic plots, including conditional weighted residual vs. observed concentrations and time (Figure 2.S2-2.S3 in *Supplementary material to Chapter 2*), population and individual predicted vs. observed concentrations (Figure 2.S4-2.S5 in *Supplementary material to Chapter 2*), VPCs (Figure 2.S6-2.S7 in *Supplementary material to Chapter 2*), and NPDE (Figure 2.S8-2.S9 in *Supplementary material to Chapter 2*) showed that the model was able to describe adequately the data.



**Figure 2.3:** Population pharmacokinetic (PK) model structure.

**Table 2.2:** Parameter estimates of the final pharmacokinetic (PK) model. RSE: residual standard error; IIV: inter-individual variability; CI: confidence interval.

<sup>a</sup>Reported as OMEGA that is the NONMEM output for IIV.

<sup>b</sup>Reported as SIGMA that is the NONMEM output for the variance of the residual error.

Parameter	Description	Unit	Population estimate (%RSE)	Bootstrap median (90% CI)	IIV <sup>a</sup> (%RSE)	Bootstrap median (90% CI)
k <sub>a</sub>	Rate of absorption	hour <sup>-1</sup>	0.956 (27.1%)	1.02 (0.51-1.40)	1.63 (33%)	1.62 (0.53-2.72)
CL	Clearance	L/hour	1.81 (8.1%)	1.77 (1.59-2.04)	0.45 (65%)	0.45 (0.15-0.76)
Q	Intercompartmental clearance	L/hour	1.85 (31.3%)	1.73 (0.86-2.84)	-	-
V <sub>2</sub>	Central volume of distribution	L	20.80 (14.6%)	21.38 (15.83-25.70)	0.412 (38%)	0.43 (0.17-0.66)
V <sub>3</sub>	Peripheral volume of distribution	L	15.10 (19%)	14.79 (10.03-20.21)	2.32 (>100%)	2.44 (89.97-94.2)
F	Bioavailability	-	0.70 (FIXED)	0.70 (FIXED)	-	-
σ <sub>PROP</sub> <sup>b</sup>	Residual error (proportional)	-	0.018 (18.9%)		-	-

## 2.6.2. Evaluation of the advantages of Bayesian estimation methods in analyzing sparse PK data

Four comparisons have been performed as described in Table 2.3. The first two comparisons demonstrate the added value of a theoretical Bayesian framework in analyzing sparse data.

**Table 2.3:** List of comparisons.

Type of sampling	N° of samples per subject	Scenario	Notes	Priors
<b>Comparison I</b>				
Protocol sampling	1	1	Parameters allometrically scaled	Weakly-informative
				Highly-informative
				No priors
<b>Comparison II</b>				
Protocol sampling	1	1	Parameters allometrically scaled	Weakly-informative  Highly-informative
		2	$CL=CL_{adult}/2$	
		3	$CL=CL_{adult}/2,$ $V2=V2_{adult}/2$	
		4	$CL=CL_{adult}/2,$ $V2=V2_{adult}/2,$ $Q=Q_{adult}/2,$ $V3=V3_{adult}/2$	
		5	Allometric exponent of CL and $Q=0.85$	
		6	Allometric exponent of CL and $Q=2/3$	
<b>Comparison III</b>				
Protocol sampling	1	1	Parameters allometrically scaled	Weakly-informative
Optimized sampling				
<b>Comparison IV</b>				
Optimized sampling	1	1	Parameters allometrically scaled	Weakly-informative
	2			
	3			
	4			

**Comparison I:** The evaluation of this comparison was guided by the convergence of the algorithm. As reported in Table 2.4, the use of priors increases the probability to obtain a successful convergence of the NLME algorithm in case of sparse sampling from 12% to 56% and 75% for weakly-informative priors and highly-informative priors, respectively. Highly-informative priors perform slightly better compared to weakly-informative priors because they introduce less uncertainty on the model parameters.

**Comparison II:** From Comparison I it is clear that if the historical data reflect the characteristics of the pediatric population correctly, highly-informative priors are better. However, this might not be the case. It is known, for example, that allometric scaling does not work for every situation, such as in the case of enzyme maturation or other physiological processes that can differ from adult to children. For these reasons, five

---

additional scenarios (Scenarios 2 to 6) were evaluated. In Scenarios 5 and 6, the concentration-time profiles of the pediatric population were simulated with a different allometric exponent compared to the one found from historical data. In Scenarios 2 to 4, more extreme situations were considered: the pediatric population was simulated using population values for clearances and/or volumes that are half of the population values of clearances/volumes identified from historical data. For scenarios 5 and 6, the two types of priors gave comparable results in terms of ratios of posterior individual estimates. For scenarios from 2 to 4, highly-informative priors led to more unprecise estimates of area under the concentration-time curve (AUC) and the maximum (or peak) concentration ( $C_{max}$ ) compared to weakly-informative priors (Figure 2.4 and Figure 2.S11, upper panels, and Table 2.S3 in *Supplementary material to Chapter 2*).

### 2.6.3. Evaluation of the impact of optimized PK designs in increasing precision for pediatric exposure extrapolation

The last two comparisons in Table 2.3 focus on how historical information can be used together with optimization techniques to guide the design of more informative trials in the pediatric population, especially in case of rare diseases.

The number of different sampling times was set to four since a larger number of design groups did not lead to further improvement of parameter precision. Consequently, four optimal sampling windows were found: the first one between 30 minutes pre-dose and the dosing time, the second one between 15 minutes and 30 minutes post-dose, the third one between 90 minutes and 150 minutes, and the last one between 225 minutes and 240 minutes (Figure 2.S10 in *Supplementary material to Chapter 2*).

For the last two comparisons, only Scenario 1 and weakly-informative priors were considered to be more conservative since there could be some maturation processes or other changes in renal/hepatic physiology that we didn't take into account in our PK model identified on historical data.

**Comparison III:** The result of this comparison (Figure 2.5 and Figure 2.S11, left lower panel, in *Supplementary material to Chapter 2*) shows that optimizing the sampling protocol when only one sample per subject is collected does not significantly improve the estimates and the probability of having a successful run (Table 2.4). The use of priors gives the necessary support to a similar extent in both cases.

**Comparison IV:** The results of this comparison show to what extent the precision of the parameter estimates increases when more samples per subject are taken, suggesting what should be the correct number of samples to balance feasibility and validity of the study (Figure 2.5 and Figure 2.S11, lower right panel, in *Supplementary material to Chapter 2*). With two samples the probability of successful run increases (Table 2.4), but the

Use of prior knowledge for the design and analysis of pediatric PK  
trials for small rare disease populations

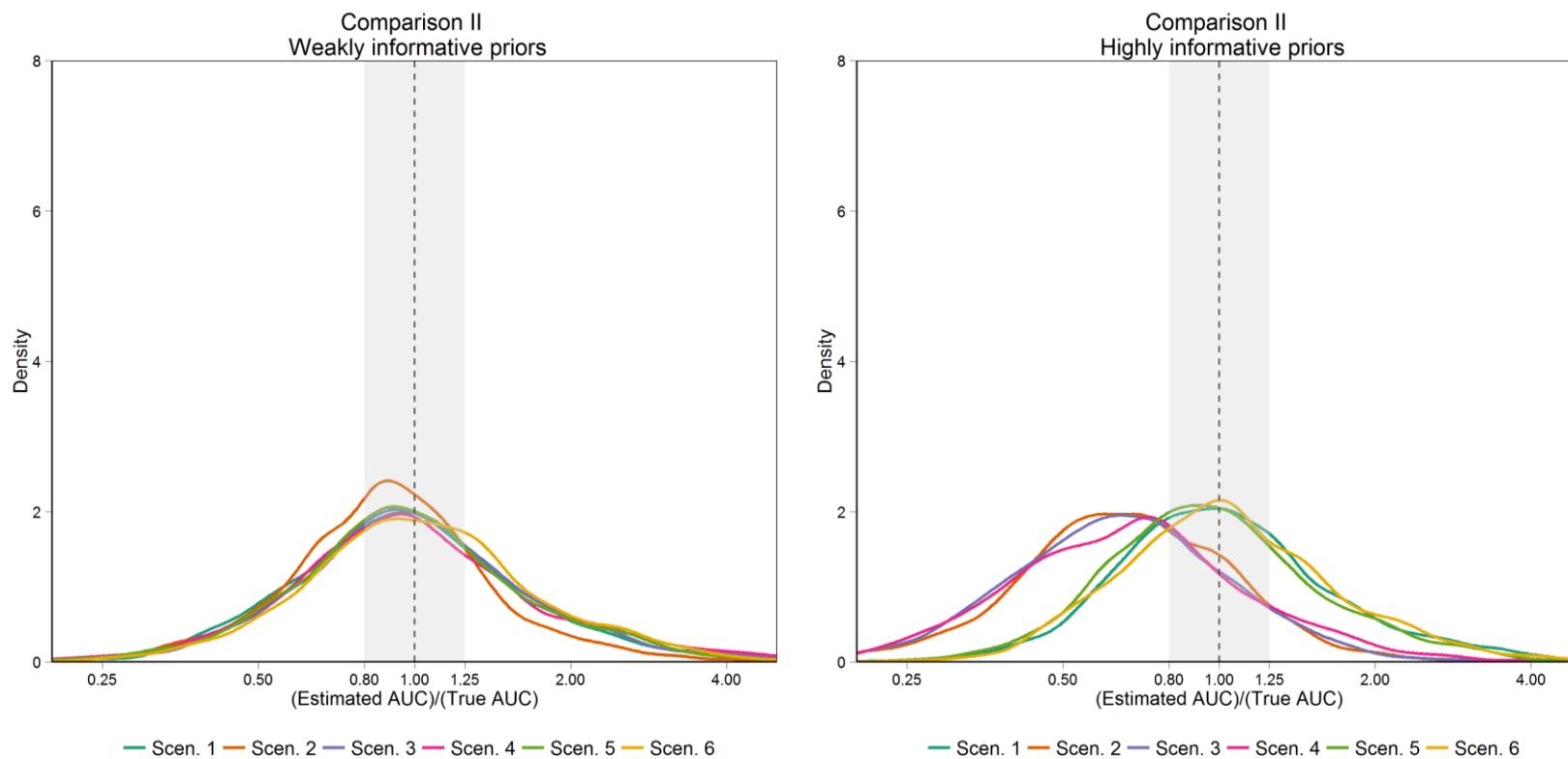
---

probability of overestimating or underestimating  $C_{\max}$  and AUC of more than 25% is still around 60% (Table 2.S3 in *Supplementary material to Chapter 2*). Increasing the number of samples to three or four shrinks the probability of having exposure-related parameter estimates outside the boundary of acceptability consistently (less than 10% with four samples) (Table 2.S3 in *Supplementary material to Chapter 2*).

**Table 2.4:** Probability of successful run for Comparison I, III, and IV. CI: confidence interval.

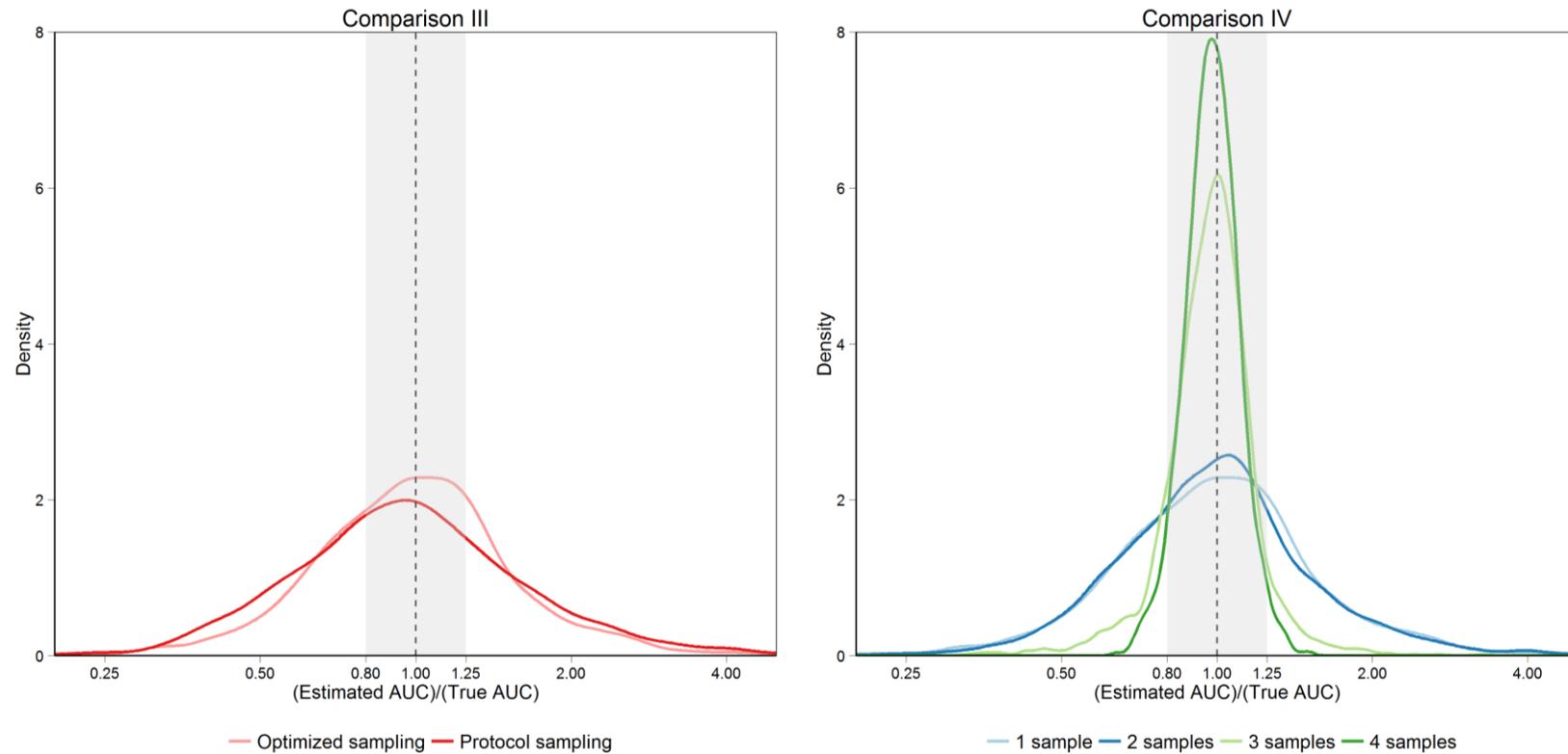
<sup>a</sup>Bootstrap median and 90% CI were calculated sampling N-times with replacement from the pool of N runs necessary to have reached 100 successful runs.

Type of sampling	N° of samples per subject	Scenario	Notes	Priors	Median of probability of successful run (90% CI)
<b>Comparison I</b>					
Protocol sampling	1	1	Parameters allometrically scaled	Weakly-informative	56.50 (50.28-62.71)
				Highly-informative	75.19 (69.17-81.20)
				No priors	12.22 (10.51-14.18)
<b>Comparison III</b>					
Protocol sampling	1	1	Parameters allometrically scaled	Weakly-informative	56.50 (50.28-62.71)
Optimized sampling					51.28 (45.64-57.43)
<b>Comparison IV</b>					
Optimized sampling	1	1	Parameters allometrically scaled	Weakly-informative	51.28 (45.64-57.43)
	2				89.96 (81.74-92.17)
	3				92.59 (88.89-93.30)
	4				94.34 (90.57-98.11)



**Figure 2.4:** Probability distributions of ratios of the estimated area under the concentration-time curve (AUC) to the original AUC (used in simulation) for Comparison II.

## Use of prior knowledge for the design and analysis of pediatric PK trials for small rare disease populations



**Figure 2.5:** Probability distributions of ratios of the estimated area under the concentration-time curve (AUC) to the original AUC (used in simulation) for Comparison III (on the left) and Comparison IV (on the right).

## 2.7. Discussion

Drugs are not usually developed for pediatric use. One of the main reasons for that is the relatively small pediatric patient population compared to adult population, which means a smaller pool of patients available for recruitment in pediatric trials. There is therefore the need to maximize the usefulness of the data obtained with the minimum number of subjects enrolled. Extrapolation of efficacy to the pediatric population from studies in adults is one of the possible ways to promote a most efficient use of data (47,48). The use of extrapolation reduces the number of trials necessary to achieve pediatric labeling but does not eliminate the need for some pediatric data. These data include, for example, PK data in the target age groups to determine the right and safe dose to be administered. Once found the dose that matches drug exposure in the pediatric population to the source population (adults), efficacy can be extrapolated with the hypothesis that similar exposure will be associated with comparable efficacy. The DEEP-2 PK sub-study is a representative case where the PK of deferasirox was investigated in a pediatric population affected by a rare disease. The use of existing data (e.g. trials in adults) as prior information can help in examining possible relationships between PK, age, and other covariates. Besides, the use of published data together with optimization techniques can provide a valuable instrument to optimize the design of new studies in the pediatric population (49,50).

This work proposes a modeling exercise to show, in a concrete case study, how prior knowledge from studies in adults can be used in data analysis in case of sparse sampling and in sampling design optimization to support pediatric-dose finding, especially in case of rare diseases. To this aim, a simulation-estimation analysis has been performed demonstrating that the use of priors always has an advantage compared to those situations where no prior information is used.

Assuming the same PK model structure, the PK parameters in the pediatric populations can be considered to have the same variability of the adult ones, but they are shifted through allometry and maturation principles. Here maturation was not considered since considerations on the ontogeny of the main metabolizing enzyme have been done. Nevertheless, it has been shown in the literature for other drugs such as paracetamol and morphine, both metabolized by UGTs, that the clearance maturation follows a sigmoid curve where the 50% of the mature value was reached at 54.6 weeks PMA for morphine and at 50.1 weeks PMA for paracetamol. A possible explanation for this delay is that UGTs are not only expressed in the liver, but also in the kidney, where they develop at the same rate as the kidney itself and seems to play an important role in the first year of life (51). Differently, in a similar work, the total body clearance was considered by one year of life 96% of the adult clearance (52). Nevertheless, the possible application of a maturation function has a limited impact on the conclusions

of this work, since, in the whole study population comprising 19 patients, it is unlikely the enrolment of neonates with less than two years in the trial.

Weakly-informative priors have proved to increase the robustness for model identifiability, allowing the integration of prior knowledge from historical data without dominating the estimation method, which is of great importance in situations where pediatric data are sparse. Nevertheless, even if weakly-informative priors are used, we are conscious that with only one sample per patient collected the intra-individual variability can be hardly distinguished from the IIV. At best, with such kind of data, good population estimates of CL/F and its IIV can be obtained, and, therefore, information on AUC. Conscious that the decision regarding the number of samples and timing of sample collected are strictly dependent on the underlying PK properties of the drug (e.g. half-life), as well as practical aspects related to implementation of PK sampling in pediatric clinical trials, with this analysis we wanted to point out the limitations of the original design and clarify expectations. We showed the possible drawbacks of collecting only one sample since they lack the necessary informative content, leading to a probability of more than 60% of over/underestimating the exposure of more than 25% (Table 2.S3).

Additionally, the advantages of using optimization techniques based on prior knowledge have been shown in the definition of the optimal sampling windows. We showed here that the reliability of PK parameters and IIV parameters are strongly dependent on the number of samples collected per individual. It was demonstrated that at least three samples per individual, obtained from pre-defined optimal sampling windows, are sufficient to provide accurate and precise estimates of almost all the PK parameter and their variabilities.

Although several methods are available in the literature for integrating prior knowledge and optimizing pediatric trial design, a general lack of awareness of clinical researchers about the proposed methodologies still exist. In the recent literature, several cases of failed completed trials have been reported (53,54), where a poor dose selection contributed to a trial failure. In (53) fixed doses across a wide range of body weights have been used; in (54) no dose response was seen in the study, and it was stated that if higher doses had been evaluated, efficacy may have been demonstrated. This reinforces the importance of performing *ad hoc* PK trials in age groups where PK cannot be reliably predicted and collecting exposure-response data for testing efficacy. In the presence of adult data, sparse sampling can be allowed to address ethical constraints. In the case illustrated, it has been demonstrated that PK information from prior adult studies can be leveraged to find an optimal sparse sampling scheme. The approach introduced can be readily generalized to other drugs for which lack of PK information in pediatrics is a potential issue. Of course, the time window recommended here was associated with specific practical constraints and to the specific drug properties of deferasirox. It is also important to keep in mind that, when this linkage between adult and pediatric is done, similar clinical response,

---

exposure-response relationship, and safety issues related to the drug must be similar between the two populations.

---

## Supplementary material to Chapter 2

---

**Table 2.S1:** Summary table of DEEP-1 and DEEP-2 studies.  
PK: pharmacokinetic; MRI: magnetic resonance imaging.

Study	Study description	Study drugs	Participant age range	N° of participants	Study endpoint(s)	Timepoints of evaluation of study endpoint(s)
<b>DEEP-1</b>	PK study of deferiprone in pediatric patients	Deferiprone (oral solution)	1 month - 6 years	30	Primary and secondary PK parameters	Samples will be collected from pre-dose up to 8 hours post-dose on a single day after drug administration. A maximum of 5 samples (post-dose; each 2 ml) will be collected per patient according a predefined sampling scheme.
<b>DEEP-2</b>	Efficacy and safety study to compare deferiprone versus deferasirox in pediatric patients	Deferiprone (oral solution) Deferasirox (dispersible tablets)	1 month - 18 years	388	Percentage of successfully chelated patients assessed by serum ferritin levels (all patients) and cardiac MRI T2* (patients above 10 years of age able to have an MRI scan without sedation)	Serum ferritin will be measured every 3 months at the central laboratory and every month at the local laboratory. Cardiac MRI T2* will be measured at month 1, 6 and 12 of treatment. Primary endpoint in terms of percentage of successfully chelated patients will be assessed as difference between basal and final (12 months) levels.
	PK sub-study of deferasirox in pediatric patients	Deferasirox (dispersible tablets)	1 year - 18 years	19	Primary and secondary PK parameters	On Visit 15 (month 12) a sample (2 mL) for the assessment of peak drug concentrations should be collected within 4 hours after dosing according to the sparse sampling scheme

						specified in the study procedure manual.
--	--	--	--	--	--	--

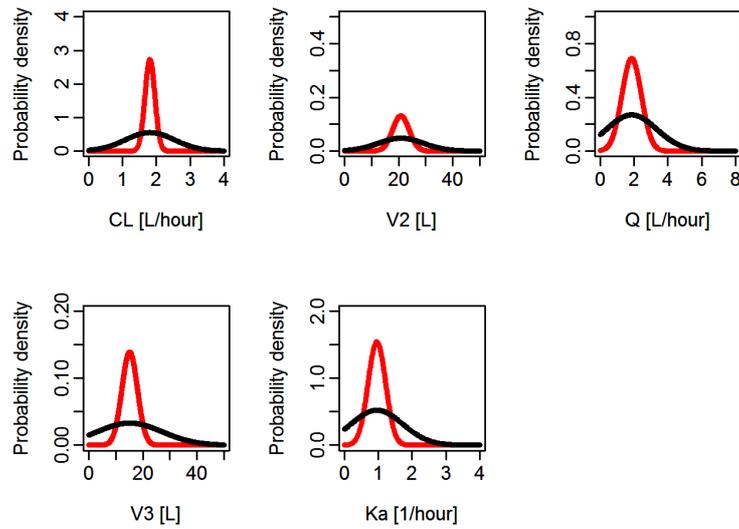
**Table 2.S2:** Visit schedule and evaluations of the DEEP-2 study. ECG: electrocardiogram; MRI: magnetic resonance imaging; CHQ: child health questionnaire.

	Run in			Treatment													13 Follow-up
	Day		Baseline Day 0	Month													
	Screening -28→-8	Washout -7 -6→-1		1	2	3	4	5	6	7	8	9	10	11	12		
Visit	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Demographic characteristics	X																
Randomization			X														
Informed consent	X																
Inclusion/exclusion criteria		X															
Pregnancy test	X															X	
Physical examination	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Medical history and current medical conditions	X																
Pharmacokinetics <sup>1</sup>																X	
Vital signs	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Liver function history	X																
Heart function history	X																
Liver MRI			X													X	
Serum Ferritin <sup>2</sup>	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
ECG	X		X		X				X			X				X	
Cardiac MRI T2*			X						X							X	
Urinalysis	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Renal function	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

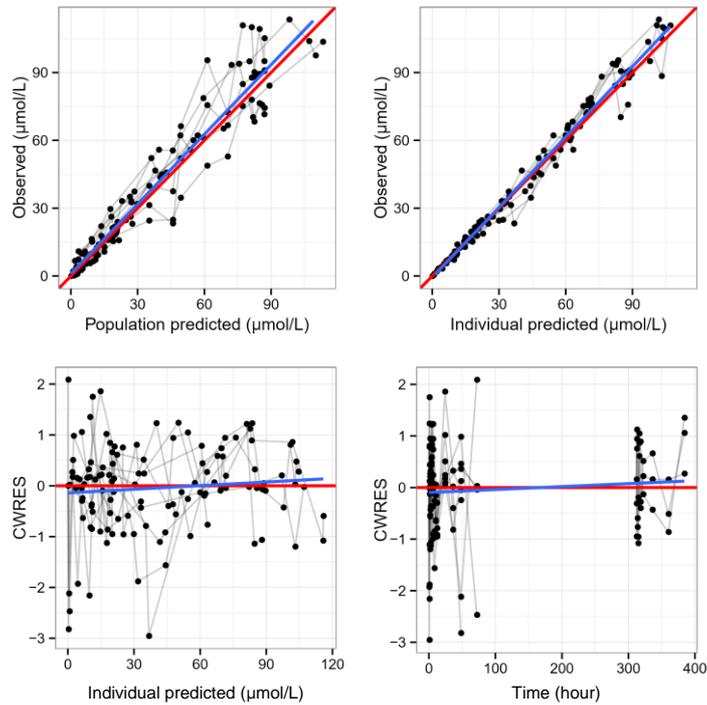
Hemoglobin	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Neutrophil count	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Hepatitis serology	X																X	
Hematology/ Biochemistry	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Ocular and audiometric test				X													X	
Concomitant medications				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Medical events				X														
Adverse events					X	X	X	X	X	X	X	X	X	X	X	X	X	X
Body height/weight	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Pubertal staging				X						X							X	
Compliance					X	X	X	X	X	X	X	X	X	X	X	X	X	
CHQ questionnaire				X						X							X	
Healthcare Resources					X	X	X	X	X	X	X	X	X	X	X	X	X	
				<p><sup>1</sup> A sample (2 mL) for the assessment of peak drug concentrations should be collected within 4 hours after dosing according to the sparse sampling scheme specified in the study procedure manual.</p> <p><sup>2</sup> Ferritin at V1, V4, V5, V7, V8, V10, V11, V13 and V14 will be evaluated only at the local laboratory. All other ferritin assessments will be evaluated at both the local and central laboratories.</p>														

**Table 2.S3:** Probability (%) of ratios between (1/1.25;1.25) for the area under concentration-time curve (AUC) and for the maximum concentration ( $C_{max}$ ) estimates for Comparison II, III, and IV.

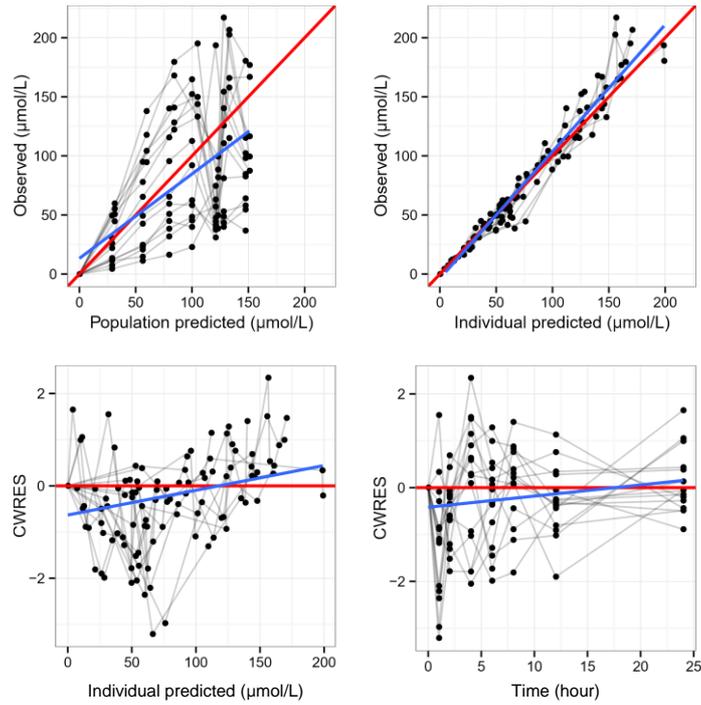
Type of sampling	N° of samples	Scenario	Notes	Priors	Probability (%) of AUC ratios between (1/1.25;1.25)	Probability (%) of $C_{max}$ ratios between (1/1.25;1.25)
<b>Comparison II (Weakly-informative priors)</b>						
Protocol sampling	1	1	Parameters allometrically scaled	Weakly-informative	37	68
		2	$CL=CL_{adult}/2$		43	72
		3	$CL=CL_{adult}/2$ , $V2=V2_{adult}/2$		37	68
		4	$CL=CL_{adult}/2$ , $V2=V2_{adult}/2$ , $Q=Q_{adult}/2$ , $V3=V3_{adult}/2$		36	66
		5	Allometric exponent of CL and $Q=0.85$		38	68
		6	Allometric exponent of CL and $Q=2/3$		37	66
<b>Comparison II (Highly-informative priors)</b>						
Protocol sampling	1	1	Parameters allometrically scaled	Highly-informative	38	70
		2	$CL=CL_{adult}/2$		26	66
		3	$CL=CL_{adult}/2$ , $V2=V2_{adult}/2$		23	41
		4	$CL=CL_{adult}/2$ , $V2=V2_{adult}/2$ , $Q=Q_{adult}/2$ , $V3=V3_{adult}/2$		23	37
		5	Allometric exponent of CL and $Q=0.85$		38	71
		6	Allometric exponent of CL and $Q=2/3$		39	68
<b>Comparison III</b>						
Protocol sampling	1	1	Parameters allometrically scaled	Weakly-informative	37	68
Optimized sampling					42	57
<b>Comparison IV</b>						
Optimized sampling	1	1	Parameters allometrically scaled	Weakly-informative	42	57
	2				46	61
	3				82	88
	4				93	94



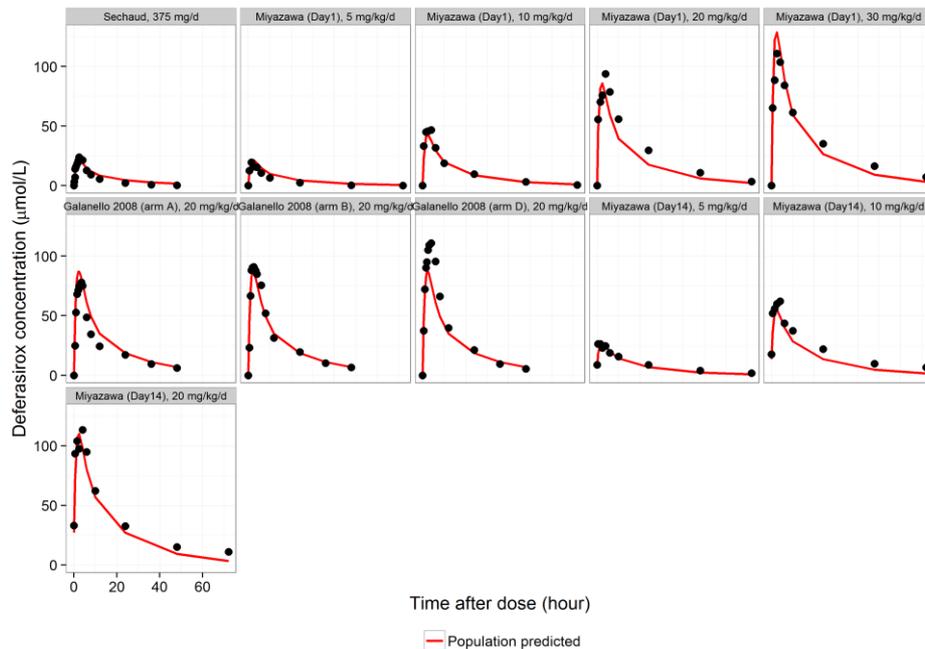
**Figure 2.S1:** Comparison between weakly-informative priors (in black) and highly-informative priors (in red).



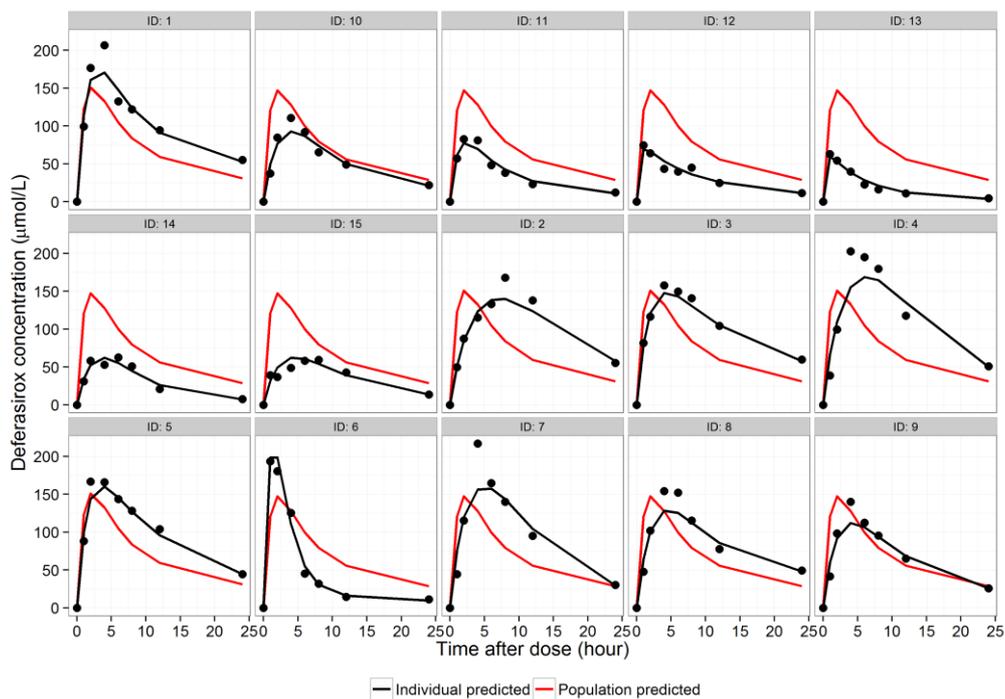
**Figure 2.S2:** Goodness-of-fit plots (GOF) for the first estimation with only mean data. Plot of observed concentrations vs. individual (top-right panel) and population predicted (top-left panel). Line of identity in red and regression lines of data points in blue. Plot of conditional weighted residuals (CWRES) vs. time (bottom-right panel) and individual predicted (bottom-left panel). Zero line in red and regression lines of data points in blue.



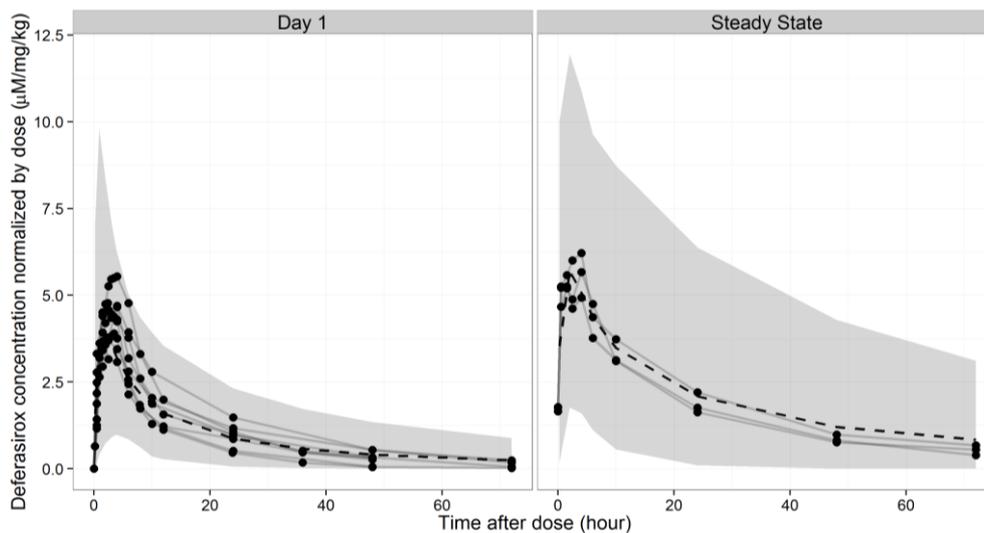
**Figure 2.S3:** Goodness-of-fit plots (GOF) for the second estimation with only individual data. Plot of observed concentrations vs. individual (top-right panel) and population predicted (top-left panel). Line of identity in red and regression lines of data points in blue. Plot of conditional weighted residuals (CWRES) vs. time (bottom-right panel) and individual predicted (bottom-left panel). Zero line in red and regression lines of data points in blue.



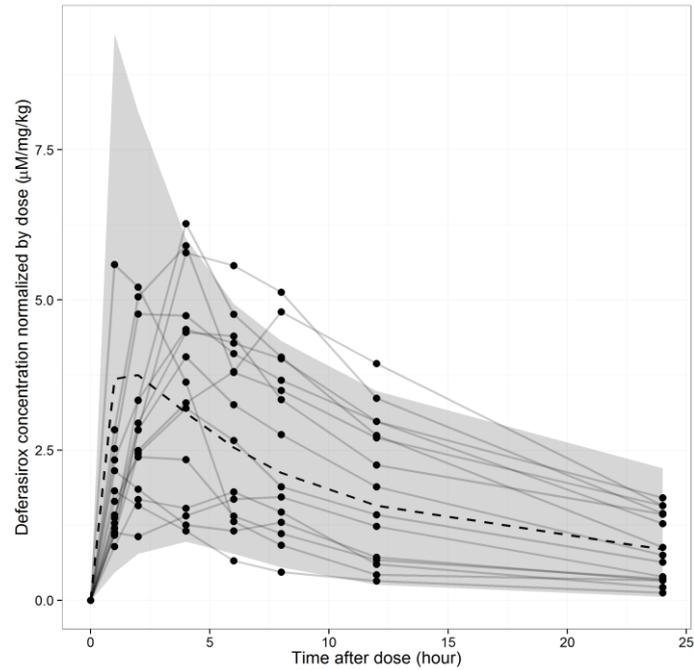
**Figure 2.S4:** Population predicted (red lines) and observed data values (black circles) vs. time after dose for the first estimation with only mean data.



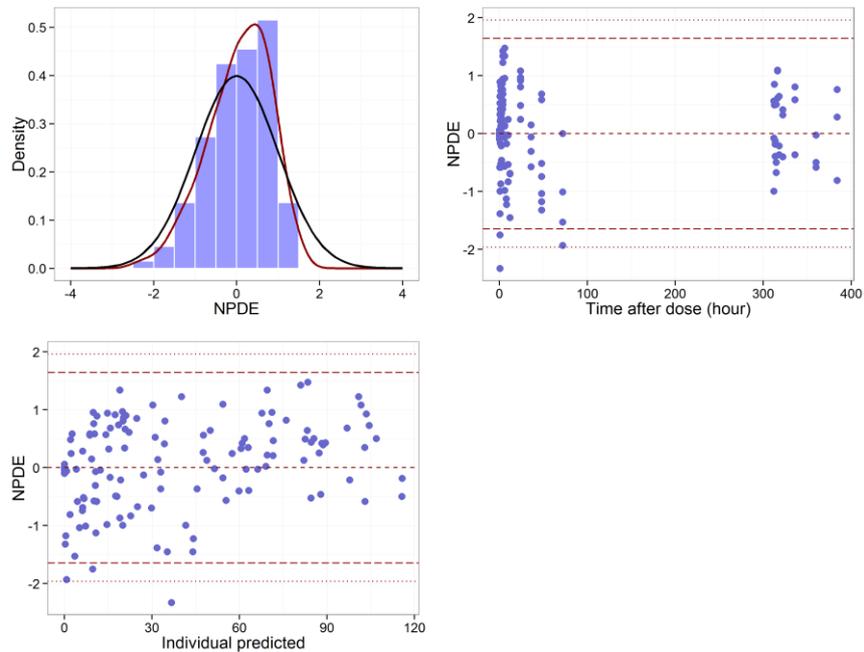
**Figure 2.S5:** Individual predicted (black lines), population predicted (red lines), and observed data values (black circles) vs. time after dose for the second estimation with only individual data.



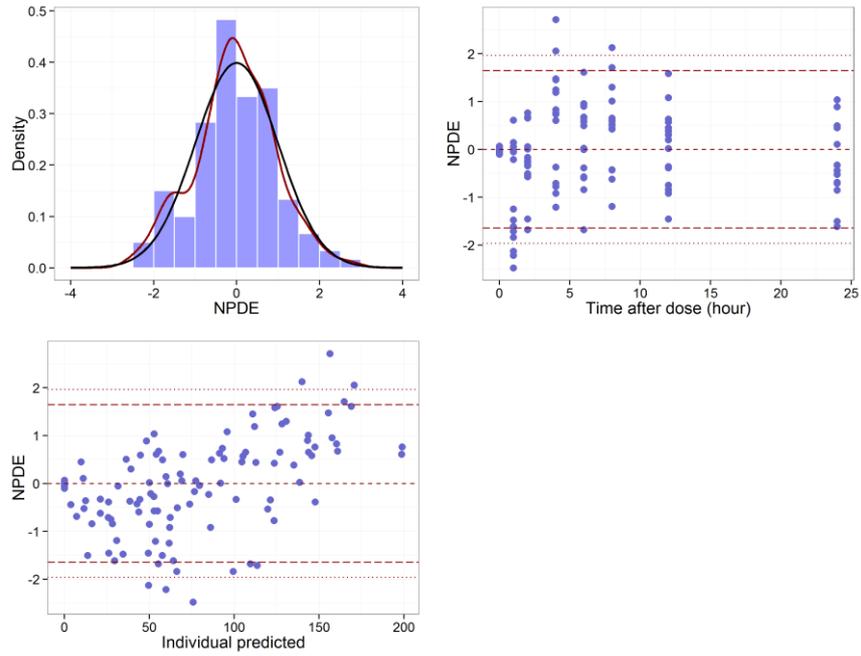
**Figure 2.S6:** Visual predicted check (VPC) plots for the first estimation with only mean data collected at Day 1 and Day 14 (at steady state). The observed data (black circles) were overlaid with predicted median (dashed black line), and 95% prediction interval (PI) (shaded grey area).



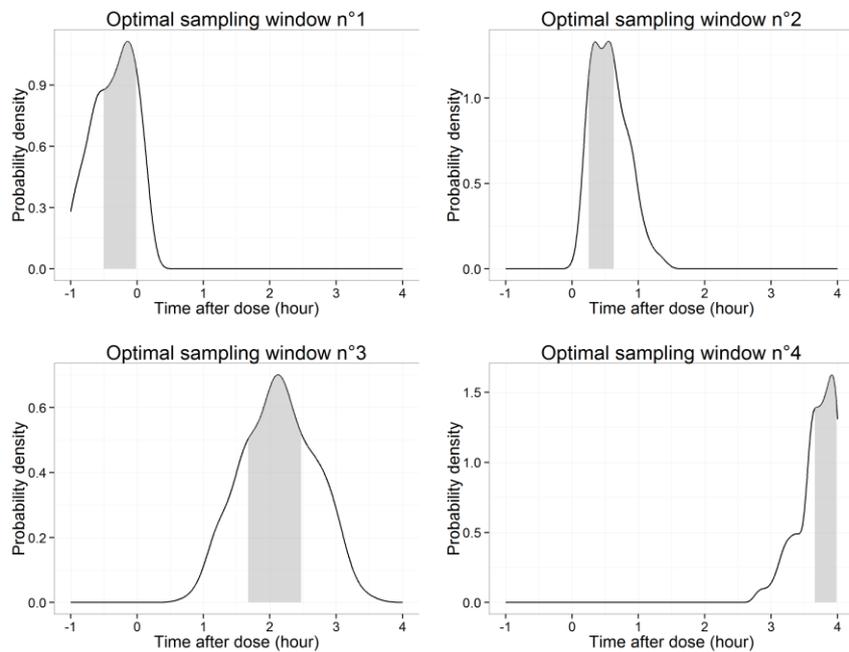
**Figure 2.S7:** Visual predicted check (VPC) plots for the second estimation with individual data. The observed data (black circles) were overlaid with predicted median (dashed black line), and 95% prediction interval (PI) (shaded grey area).



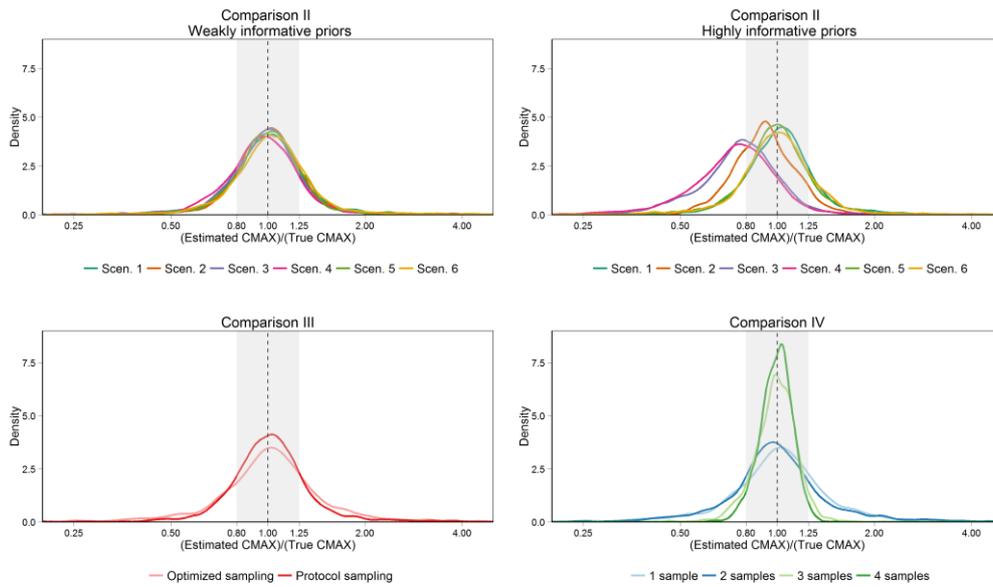
**Figure 2.S8:** Histogram of normalized predictive distribution error (NPDE) with the density of the standard distribution overlaid (top-left panel), scatter plot of NPDE vs. time after dose (top-right panel), scatter plot of NPDE vs. population predicted concentration (bottom panel) for the first estimation with mean data.



**Figure 2.S9:** Histogram of the normalized predictive distribution error (NPDE) with the density of the standard distribution overplayed (top-left panel), scatter plot of NPDE vs. time after dose (top-right panel), scatter plot of NPDE vs. individual predicted concentration (bottom panel) for the second estimation with individual data.



**Figure 2.S10:** Probability density of selected optimal sampling times for each of the four design groups (black curves) and optimal sampling windows (grey shaded areas).



**Figure 2.S11:** Probability distributions of ratios of the estimated maximum concentration ( $C_{\max}$ ) to the true  $C_{\max}$  (used in simulation). Upper panels refer to Comparison II, lower left panel to Comparison III and lower right panel to Comparison IV.

---

# Chapter 3

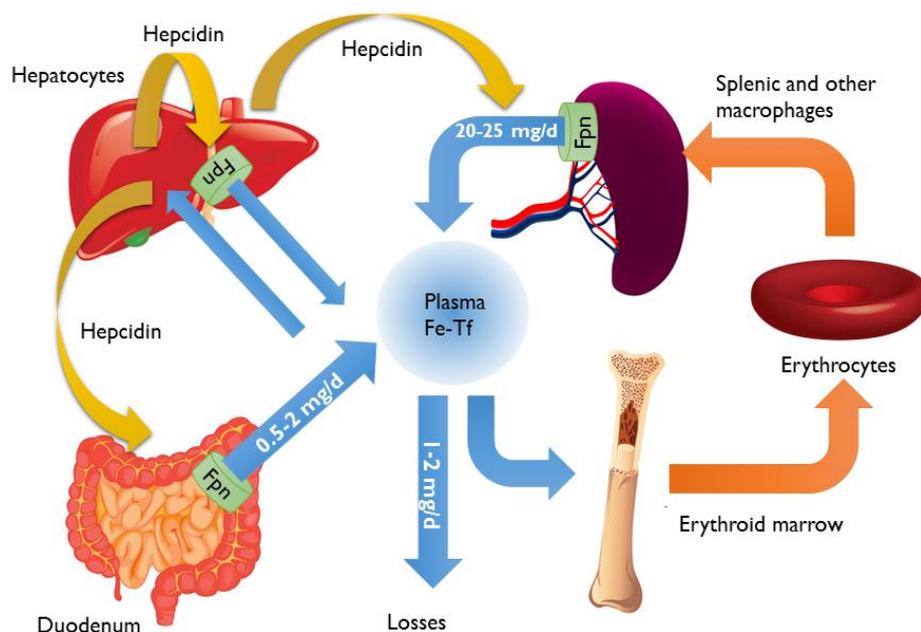
---

## Model-based predictions, including exploiting prior knowledge, of patient-response to chelation therapy

### 3.1. Introduction

Iron overload is one the most relevant complications occurring in patients affected by rare hemoglobinopathies (e.g. beta-thalassemia), which requires sustained transfusions to survive, and it results from defects in the regulation of iron balance.

The total body iron content in healthy adults is usually about 3-5 g of iron (45-60 mg/kg body weight) (55). Only 1-2 mg of dietary iron is daily absorbed in the intestine and then is lost due to menstrual blood, bleeding, sweating, skin desquamation, and urinary excretion (56) (Figure 3.1). About two third of the total iron content is found in hemoglobin (Hb) in red blood cells (RBC) (55); 20-25 mg of iron is recycled from aging erythrocytes by reticuloendothelial macrophages in the spleen or other organs for new RBC synthesis (57). Transferrin is the protein responsible for transporting iron in plasma to most cells of the body. Only around 0.1% of the total iron content is found in this transit compartment (56). Stored iron is principally located in the hepatocytes in the liver, where, in a typical adult man, from 0.5 to 1 g of iron are stored (56,58). Hepatocytes can incorporate both transferrin-bound and free iron circulating in the plasma, even if free iron is found only in iron-overloaded patients. Inside the hepatocytes, iron is found in ferritin, an iron storage protein. Ferritin forms a roughly spherical container, capable of holding up to 4,500 iron atoms (59). The normal range for ferritin is very large (18-350 ng/ml), and it varies between men and women, and with age (60).



**Figure 3.1:** Distribution of iron in the body and its regulation by hepcidin and ferroportin (Fpn). Iron bounded to transferrin (Fe-Tf) is depicted in blue, while iron in erythrocytes in orange. Adapted from *Ganz T et al., Physiol Rev, 2013 (57)*.

In case of need, hepatocytes can export iron, which returns to the systemic circulation. Reticuloendothelial macrophages of the liver and spleen are not only essential because they are responsible for recycling iron from senescent RBC, but also because they can also store it. Similar to hepatocytes, macrophages have a mechanism to export iron in case of need (56) (Figure 3.1).

Iron homeostasis is regulated mainly by the hormone hepcidin and the cellular iron exporter ferroportin (Fpn), together with iron regulatory proteins that bind iron-responsive elements in regulated mRNAs, and hypoxia-inducible factors that control the transcription of genes involved in the maintenance of iron metabolism (55). Hepcidin is synthesized in hepatocytes and then circulates in plasma. It regulates the expression of the iron exporter Fpn on the surface of enterocytes, macrophages, and hepatocytes through internalization followed by degradation (Figure 3.1). The consequence is that less iron is exported from the intestine and the iron stores in hepatocytes and macrophages (55).

Blood transfusions in thalassemia major patients induce an increased exposure of iron from macrophages (each unit of packed RBC contains 200-250 mg of iron), resulting in a saturation of transferrin after a long period (61). This has as a consequence the presence of non-transferrin-bound iron in the plasma which is taken up by organ tissues. Non-transferrin-bound iron promotes the generation of free radicals, causing tissue damage, and an

overwhelming of ferritin storage capacity, which in turn induces ferritin degradation by lysosomes and the consequent formation of an insoluble iron complex, i.e. hemosiderin, leading to organ toxicity.

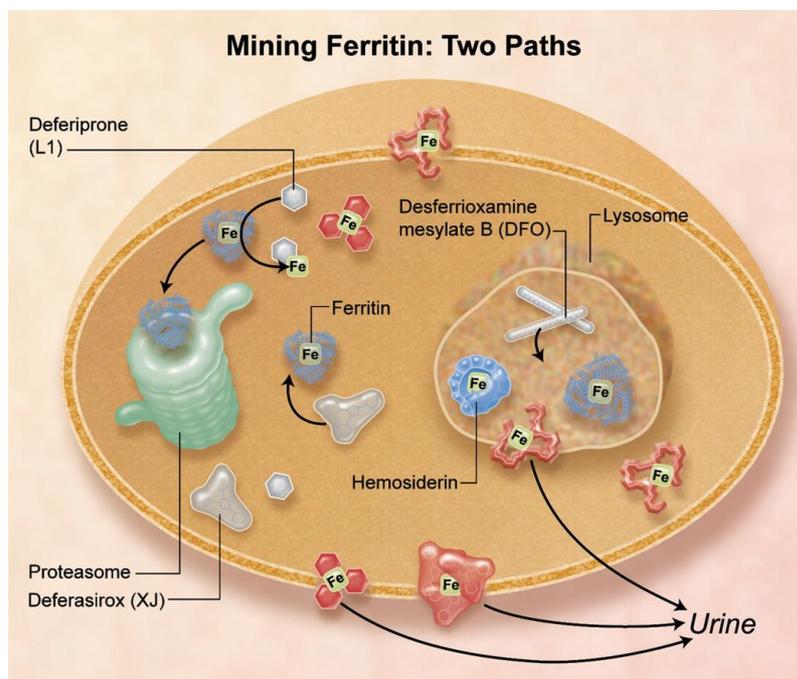
Therapy with chelating agents is required in case of iron overload to promote iron extraction, clearing plasma from non-transferrin-bound iron, removing iron from cells, and restoring body iron content to normal levels.

Iron chelators are required to facilitate iron excretion through either the urine and feces since the body does not possess any active internal mechanism to remove excess iron from sustained transfusions.

Three are three iron-chelating agents available: deferoxamine, deferiprone, and deferasirox (62).

Deferoxamine has been available for more than three decades, and it is considered the first-line treatment for iron overload. It is poorly absorbed from the intestine, and rapidly eliminated in urine and feces (half-life of 20 minutes); therefore, it must be given by intravenous or subcutaneous continuous infusion (63). Deferoxamine exerts its action entering cells by endocytosis, where it induces ferritin entry into lysosomes. The iron released from lysosomal degradation of cytosolic ferritin is then bound by deferoxamine, and the chelated iron can leave in this way the cell (64) (Figure 3.2). Several dose-dependent adverse events, such as audiometric, retinopathic, and growth effects, limits its use and leads to poor compliance.

Differently from deferoxamine, deferiprone and deferasirox target cytosolic iron, thus preventing its incorporation with ferritin (Figure 3.2).



**Figure 3.2:** Iron chelators and their pathways. Adapted from *Theil EC, Blood, 2009 (65)*. L1: Deferiprone; XJ: Deferasirox; DFO: Deferoxamine.

---

Deferiprone is orally active and is more efficient in preventing cardiac damage, especially when is given in combination with deferoxamine (66). It is characterized by a short half-life and by rapid inactivation by glucuronidation; for these reasons, it is usually administered three times daily, which has an impact on patient compliance. Nevertheless, it could be associated with neutropenia and/or agranulocytosis, demanding frequent neutrophil counts to monitor patient status.

Deferasirox is a tridentate iron chelator, well absorbed from the gut and slowly cleared from the circulation (long half-life), which allow a once-daily oral administration. Deferasirox is mainly metabolized by glucuronidation and only less than 10% by oxidative metabolism by CYPs. Deferasirox and its metabolites are then eliminated in feces.

Deferiprone and deferasirox are the most recent iron chelators appeared on the market, and they have been extensively studied in adults, where they both proved to have significant benefits regarding life quality improvements compared to the oldest therapy deferoxamine. Despite that, still limited evidence on their use in the pediatric population is available. Because of this, the DEEP-2 study (hereafter called also DEEP-2 non-inferiority study) was implemented to evaluate the non-inferiority of deferiprone to deferasirox in pediatric patients affected by hereditary hemoglobinopathies, further providing additional efficacy and safety data after one-year treatment with these two iron chelators in the target population (Table 2.S1 in *Supplementary material to Chapter 2*).

Given the considerable evidence present in the literature (67–70) that serum ferritin is a reliable parameter to evaluate chelation efficacy, in the DEEP-2 non-inferiority study the percentage of successfully chelated patients was based on ferritin measurements at baseline and at the end of the trial. Additionally, serum ferritin levels were measured in the entire population every month at local laboratories and every 3 months both at local and the central laboratories (Table 2.S2 in *Supplementary material to Chapter 2*).

Serial measurement of serum ferritin at regular intervals are in fact recommended by several clinical practice guidelines (67,68), which state that not only the interpretation of absolute serum ferritin values, but also the interpretation of trends is necessary to monitor iron chelation efficacy, and, consequently, to adjust or modify chelation regimens when required (69). Nevertheless, few pieces of evidence regarding the predictive utility of serum ferritin trends over time has been reported. A possible drawback could be interpreting as a lack of response the absence of a serum ferritin decrease in the first few months of chelation regimen; in fact, the absence of a decreasing trend cannot exclude a decreasing iron burden (70).

Therefore, it is important to understand the utility and limitations of serial serum ferritin measurements in the prediction of response to chelation regimens. To this aim, first, a PK-PD model for iron overload from available prior knowledge has been successfully developed. Then, using the DEEP-2 non-inferiority study as a case study, we aimed to demonstrate how serum ferritin trends together with the newly developed PK-PD model can be used

---

to predict clinical response in pediatric patients treated with deferiprone or deferasirox, and, therefore, to optimize their drug therapy, reducing iron burden in these patients.

## **3.2. Development of a PK-PD model for iron overload and chelation therapy**

The first aim of the work presented in this Chapter is the development of a PK-PD model that was able to describe iron dynamics in patients affected by iron overload. Serum ferritin was used as the clinical biomarker to account for disease progression and drug effect.

### **3.2.1. Clinical efficacy data**

Since no clinical data were available at the beginning of this work, a literature research was performed to find all relevant publications. Serum ferritin data were extracted from several published clinical studies (71–83) (Figure 3.S1). Serum ferritin profiles in untreated patients at first, followed by profiles of treated patients were pooled together. All the studies considered involved patients affected by transfusion-dependent hemoglobinopathies, especially beta-thalassemia major, to which deferasirox or deferiprone was administered as monotherapy. Details on patient demographics, study protocols, and type of data (mean or individual data) are described in Table 3.1. When information about patient covariates of interest was missing, assumptions had to be made. A detailed description of all the assumptions undertaken is provided in *Supplementary material to Chapter 3*.

**Table 3.1:** Efficacy studies considered for the analysis.  
L1: deferiprone; XJ: deferasirox.

Reference	Control arm?	Drug	Dose (mg/kg/d)	N° subjects	Individual data?	Age (yrs)
(75)	No	L1	75	20	Yes	33.27
(72)	No	L1	78.2	12	No	15.9
(76)	No	L1	From 75 to 100 in 8 weeks	29	No	25.1
(73)	No	L1	75	60 65 26	No	21.3
(83)	No	L1	Individual dose reported	8	Yes	Not available
(77)	No	XJ	6.2 10.2 19.4 28.2	15 78 84 119	No	17
(78)	No	XJ	From 50 to 100	609 984 150	No	30.6
(71)	Yes	-	-	24	Yes	Individual age reported
(74)	Yes (one arm)	L1	50 75 0	30 21 24	No	4-14
(79)	No	XJ	11.3	20 20	No	6.7 14.1
(80)	No	XJ	10 20	24 24	No	23.7 25.6
(81)	No	L1	75	71	Yes	20
(84)	No	L1	75	20	Yes	Not available

### 3.2.2. PK modelling

A two-compartment PK model with first-order absorption and elimination was used to describe the time-course of deferasirox plasma concentration (see *Chapter 2*). A one-compartment model with first-order absorption was used to describe the PK of deferiprone as reported in (85).

The steady-state average concentration ( $C_{SS}^{AV}$ ) was then derived for the population of interest with the following formula (Eq. 3.1):

$$C_{SS}^{AV} = \frac{DOSE \left[ \frac{mg}{kg} \right] \cdot Weight[kg]}{CL/F_{scaled}} \quad (3.1)$$

Body weight was assigned as a covariate for clearance and volume. An allometric scaling with a fixed exponent of 0.75 and a reference weight of 70 kg was applied to scale clearance from adults to the pediatric population. The final PK parameters used to calculate the  $C_{ss}^{AV}$  include the estimated apparent clearance (CL/F) of 2.58 L/h for deferasirox and of 30.8 L/h for deferiprone (85).

### 3.2.3. PK-PD modeling

A PK-PD model for iron chelation in chronic-transfused patients with iron overload was developed starting from literature knowledge of iron metabolism.

It consisted in a single compartment representing the surplus of iron stored in the whole body. Since the natural intake of iron (from the diet) is negligible compared to the iron intake due to blood transfusions (0.5-2 mg of daily dietary iron vs. 200-250 mg from a single transfusion), only a zero-order input rate, correlated to the annual blood consumption per unit of body weight, was considered (Eqs. 3.2 and 3.4). A conversion factor of 1.16 mg/ml was applied to the annual blood consumption since it is known from the literature that 100 ml of RBC per kg body weight corresponds to 116 mg of iron per kg body weight (70) (Eq. 3.4).

The surplus of iron is then linked to the ferritin measured in plasma through an Emax relationship to account for a saturation effect for high iron values (Eq. 3.3); a linear relationship was also tested.

Without chelation therapy, the elimination from the iron compartment is a zero-order because the human body doesn't have any active process to remove iron in excess, but only a constant small amount of iron (0.5-2 mg) is eliminated daily due to bleeding, sweating, skin desquamation, and other processes. This elimination can be considered negligible compared to the drug effect, and, therefore, was not considered in the model. The drug effect was described as a first-order elimination from the iron compartment (Eqs. 3.2 and 3.5). Different concentration-effect relationships (linear model, Emax model, Hill model) have been tested to characterize the drug effect. Drug  $C_{ss}^{AV}$  was used as a measure of drug exposure (Eq. 3.5). The final model is therefore described by the following equations:

$$\frac{dIRON(t)}{dt} = K_{in} - K_{out} \cdot IRON(t), \quad (3.2)$$

$$IRON(t = 0) = IRON_0,$$

$$FERRITIN(t) = \frac{FerMax \cdot IRON(t)}{Fer50 + IRON(t)},$$

(or  $FERRITIN(t) = SlopeFer \cdot IRON(t)$ ), (3.3)

$$K_{in} = 1.16 \cdot BLOODCONS, \quad (3.4)$$

$$K_{out} = f(C_{ss}^{AV}), \quad (3.5)$$

where  $BLOODCONS$  is the RBC consumption per month (i.e. annual blood consumption/12), and  $f(Css^{AV})$  is the concentration-effect relationship. The initial iron content  $IRON_0$  is derived from ferritin baseline ( $BASELINE_{ferritin}$ ), reversing the relationship between ferritin and iron.

Fixed and random effects were introduced in a stepwise manner. Inter-individual variability (IIV) and inter-study variability (ISV) was assumed to be log-normally distributed.

Model building and evaluation were performed using NONMEM v.7.3 (Icon Development Solution, USA). Conditional estimation with interaction was used as estimation method.

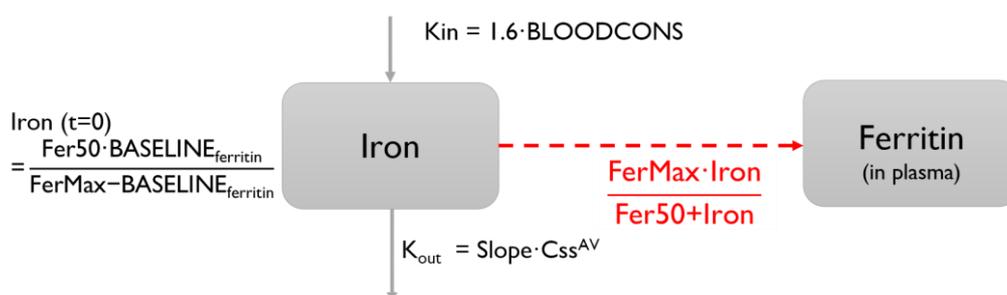
Selection of the best model was based on objective function values, completion of estimation and covariance steps, precision of the parameter and error estimates, number of significant digits, correlation between parameters, and absence of zero gradients.

A visual inspection of GOF plots was used to assess fitting performances. These include: population and individual vs. observed ferritin concentrations, and CWRES vs. observed ferritin concentration (or time). The final model validation was based on VPC and NPDE. R v.3.0.3 was used for GOF plots, NPDE, and VPC.

Bootstrap (1,000 samples) was used to evaluate the accuracy of parameter estimates (standard error and confidence intervals). The bootstrap was performed in PsN v.4.2.

### 3.2.4. Results

A schematic representation of the final PK-PD model is represented in Figure 3.3.



**Figure 3.3:** Schematic representation of the pharmacokinetic-pharmacodynamic (PK-PD) model.

Ferritin turnover was described by an Emax model, linking the surplus of iron to the presence of plasma ferritin. The impact of blood transfusions is indirectly linked to ferritin since it is correlated with the input rate to the

compartment (Eqs. 3.2-3.3). The baseline serum ferritin is used to calculate the initial iron content at  $t = 0$  as follows:

$$IRON(t = 0) = \frac{Fer50 \cdot BASELINE_{ferritin}}{FerMax - BASELINE_{ferritin}}. \quad (3.6)$$

A significant improvement in the model was provided by adding an IIV on the  $K_{in}$  parameter (applied only for those studies where individual data were available). Since the baseline serum ferritin value could be affected by measurement error, the residual error on this term was subtracted as follows (Eq. 3.7):

$$BASELINE_{ferritin} = \frac{BASELINE_{ferritin_{COV}}}{e^{(\eta_{BASELINE_{COV}})}} \quad (3.7)$$

where  $BASELINE_{ferritin_{COV}}$  is the value of the covariate as reported in the published clinical study, and  $\eta_{BASELINE_{COV}}$  is the residual error term on  $BASELINE_{ferritin_{COV}}$ , which is normally distributed with mean 0 and variance to be estimated.

The effect of both deferasirox and deferiprone was introduced as enhancing the natural degradation rate of serum ferritin. A linear model was considered since it accurately describes changes in serum ferritin, which implies  $K_{out} = SlopeDrug \cdot Css^{AV}$ , where two different  $SlopeDrug$  parameters were estimated for deferiprone and deferasirox. This model also takes into account the fact that there is no elimination before treatment. Non-linear models (Emax and Hill models) were also tested but the data available (characterized by limited dose ranges) did not allow the estimation of their parameters. IIV and ISV on the drug effect parameter of deferiprone consistently improved model fitting performances. The absence of individual data and the limited number of studies involving deferasirox did not support the estimation of IIV and/or ISV on the corresponding drug effect parameter. A proportional error model was used to describe residual error variability. The final model parameters and the bootstrap results are presented in Table 3.2.

Model diagnostics, such as GOF plots, VPC (Figures 3.4-3.5 and Figures 3.S2-3.S3), and NPDE (not reported here) reveals the capability of the model in describing well the data.

**Table 3.2:** Parameter estimates of the final pharmacokinetic-pharmacodynamic (PK-PD) model.

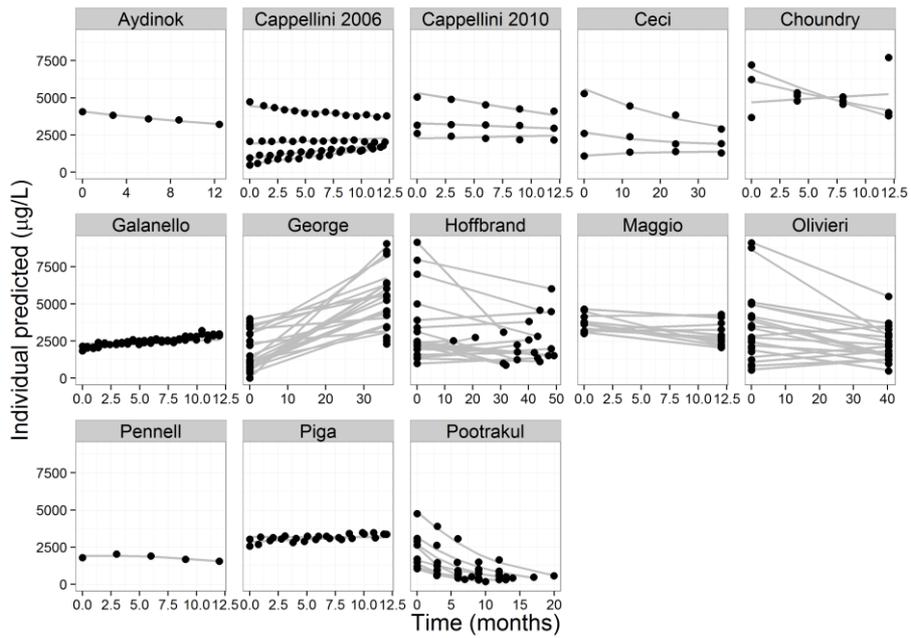
IIV: inter-individual variability; ISV: inter-study variability; RSE: residual standard error; CI: confidence interval; L1: Deferiprone; XJ: Deferasirox.

<sup>a</sup>Reported as OMEGA(N) that is the NONMEM output for IIV.

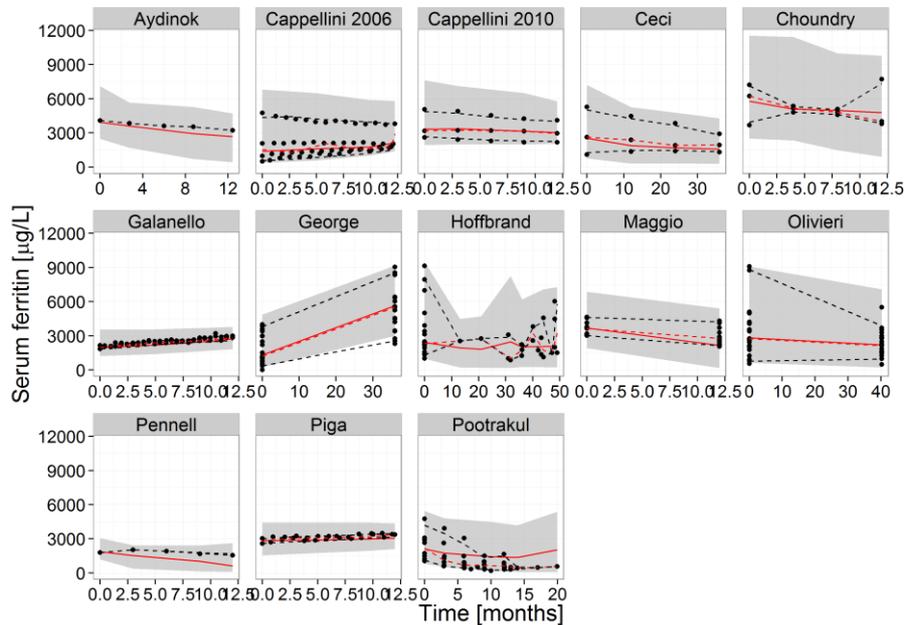
<sup>b</sup>Reported as SIGMA(N) that is the NONMEM output for the variance of the residual error ( $\sigma_{PROP}^2$ ).

<sup>c</sup>Reported as OMEGA(N) that is the NONMEM output for ISV.

Parameter	Unit	Population estimate (%RSE)	Bootstrap median (90%CI)
FerMax	-	13.6 · 10 <sup>3</sup> (9.4%)	15.10 · 10 <sup>3</sup> (10.41 · 10 <sup>3</sup> -1,317.1 · 10 <sup>3</sup> )
Fer50	[iron]	2,030 (12.8%)	2,317 (1,383-293,210)
Slope (L1)	(month · [conc]) <sup>-1</sup>	0.0109 (3.1%)	0.0107 (0.0075-0.0135)
Slope (XJ)	(month · [conc]) <sup>-1</sup>	0.0013 (8.3%)	0.0013 (0.0007-0.0018)
IIV on Slope (L1) <sup>a</sup>		0.244 (58%)	0.2064 (0.1266-0.3178)
ISV on Slope (L1) <sup>c</sup>		1.17 (39%)	1.2006 (0.8381-1.6671)
IIV on <i>BASELINE</i> <sub>ferritin<sub>COV</sub><sup>1</sup></sub>		0.318 (3.2%)	0.3109 (0.2064-0.4006)
IIV on <i>BLOODCONS</i> <sup>1</sup>		0.418 (23%)	0.3896 (0.1958-0.7045)
$\sigma_{PROP}$ <sup>b</sup>		0.00773 (4%)	0.00677 (0.00397-0.01141)



**Figure 3.4:** Plots of observed serum ferritin concentrations (black circles) and individual model predictions (grey lines) for the final model pharmacokinetic-pharmacodynamic (PK-PD) model.



**Figure 3.5:** Visual predictive check (VPC) plots of the final pharmacokinetic-pharmacodynamic (PK-PD) model stratified by study. The observed data (black circles) were overlaid with predicted median (red line), median of observed data (dashed red line), 5<sup>th</sup> and 95<sup>th</sup> percentiles of observed data (dashed black lines), and 95% prediction interval (PI) (shaded grey area).

### 3.3. A model-based optimization of chelation therapy

A continuous dose adjustment together with a close monitoring are the foundations of an optimal chelation therapy in children. In general, pediatric patients require higher transfusional rate compared to adults to maintain adequate Hb levels for a normal growth and development. Consequently, intensive chelation strategies in children must be pursued to achieve the goals of chelation therapy (86).

Before adjusting the iron chelator dose in a pediatric patient, a detailed characterization of the underlying disorder is required; thus, an evaluation of all the documentation on the history of transfusion and chelation must be done. Besides, measurements of liver iron content (LIC) and serum ferritin concentrations are necessary to drive the right conclusions. Considerations on the rate of transfusional iron loading and cardiac iron deposition need also to be done (64,87).

Despite increased knowledge, several uncertainties exist with regard to the optimal approach to iron-chelating therapy, for example for the choice of the starting dose. In (78) the authors considered the baseline iron burden and the ongoing transfusional iron intake to select the dose of deferasirox; in addition, they used serum ferritin measurements to guide dose titration.

Uncertainties also exist on the best frequency to choose for serial serum ferritin measurements. One of the most commonly adopted frequency is every three months, but many guidelines do not indicate a specific one to adopt (88). In the DEEP-2 study, dose adjustments were evaluated every three months during the visits at the central laboratory (Table 2.S2 in *Supplementary material to Chapter 2*). During these visits, clinicians evaluate the possibility of a dose adjustment based on a comparison between the serum ferritin level measured during the visit and the baseline measure.

For these reasons, the principal aim of the work presented in this second section of *Chapter 3* is to investigate if serum ferritin trends together with an algorithm based on model predictions are able to increase the probability of having an acceptable chelation in pediatric patients with hemosiderosis. The concept of a model-based dose adjustment approach was first introduced by Sheiner (89). The main idea is to individualize a population PK-PD model relating dose to outcome, using data from the patient's previous responses to the drug. The dosage adjustment is then determined from individual response predictions given by the patient-specific model. The PK-PD model combined with models for covariate distribution in the target population and the DEEP-2 non-inferiority study design, allow CTS of DEEP-2 design to test the aforementioned hypothesis.

A model-based dosing algorithm adjustment strategy will be presented in the next sections and compared to the original strategy adopted in the DEEP-2 non-inferiority study. Besides, a quantitative analysis was performed to examine the impact on each strategy of the frequency of efficacy

---

assessments, which were done quarterly (i.e. every three months) in the DEEP-2 study during the visits at the central laboratory.

### 3.3.1. Simulation of a virtual pediatric population

The 388 patients enrolled in the DEEP-2 non-inferiority study are adherent to the following inclusion criteria:

- are aged from 1 month up to less than 18 years old;
- are affected by any hereditary hemoglobinopathies requiring chronic transfusion therapy and chelation;
- receive at least 150 ml/kg of packed RBC/year;
- have serum ferritin  $\geq 800$  ng/ml at screening.

Once enrolled in the study, they have been 1:1 randomized in two groups: 194 patients are administered with deferiprone (experimental arm) at 75-100 mg/kg/day for seven days per week, the other 194 patients are administered with deferasirox (standard arm) at 20-40 mg/kg/day. Deferiprone daily dose cannot exceed 100 mg/kg, while deferasirox daily dose cannot exceed 40 mg/kg.

The PK-PD model previously described include a covariate effect on multiple parameters to describe variability in the data. Therefore, a virtual patient population with representative covariate distribution is pivotal for CTS. When available, characteristics of the original population have been used.

Therefore, the blood consumption was randomly extracted from a uniform distribution between 150 and 200 ml/kg/year; the lower limit was imposed by the inclusion criteria defined in the protocol, while the upper limit was chosen according to maximum recommended rate of iron loading reported in (70).

The baseline serum ferritin was randomly extracted from a uniform distribution between 800 and 6000 ng/ml. The lower limit was imposed by the inclusion criteria, while the upper limit was chosen according to expert's opinion since these patients have been all treated since the diagnosis, values higher than 6000 ng/ml are not considered feasible.

The body weight (necessary to compute the scaled clearance and, hence,  $C_{ss}^{AV}$ ) was derived from PMA and sex with an appropriate demographic model (44). A 1:1 sex ratio was considered, and PMA was randomly extracted from a uniform distribution between 1 month and 18 years old (plus a gestational age of 40 weeks).

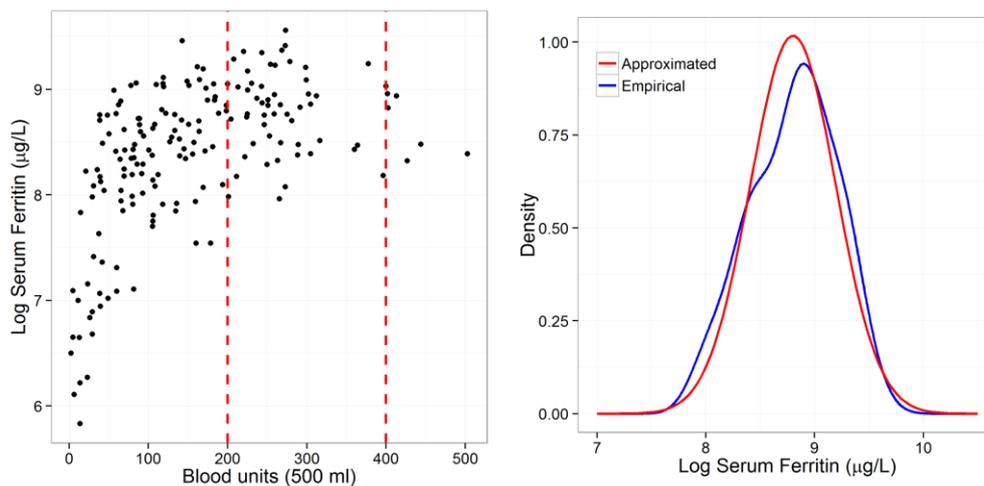
### 3.3.2. Population PK-PD model

The following changes have been done made the model originally estimated on literature data and described in paragraph 3.2.3. Since understanding variability in drug response is very important when *in vivo*

mimicking clinical trials are simulated, IIV was incorporated to avoid overly optimistic confidence in the simulation results. In particular, IIV was added on the following model parameters: *FerMax*, which defines the maximum achievable level of plasma ferritin in response to a surplus of iron, and *SlopeDrug* of deferasirox, which represents the inhibitory effect by deferasirox on body iron content.

The IIV on *FerMax* was derived from literature data of serum ferritin vs. cumulative amount of blood units (71,90,91), which were normalized to a volume of 500 ml per blood unit to ensure that equal volume of blood per transfusions was taken into account for the entire pooled dataset (Figure 3.6). The serum ferritin level seems to step increase when the first 50-100 units are given, after which further transfusions contribute only to a proportionately smaller increase. Given the strong linear correlation between liver iron content and blood units (92), the distribution of the log-transformed serum ferritin data in this saturation phase (between 200 and 400 units) was approximated with a normal distribution (Figure 3.6), thus the variance of this empirical distribution was derived and used as IIV for *FerMax*.

The IIV and ISV on *SlopeDrug* were assumed to be the same in deferasirox and defेरिप्रोन since the two drugs share the same mechanism of action (93). This modified version of the PK-PD model was used to simulate serum ferritin data every month from 0 to 12 months for 388 pediatric patients meeting the inclusion criteria of the protocol.



**Figure 3.6:** Pooled literature data of serum ferritin vs. cumulative blood units from untreated patients in semi-log scale (left panel). Empirical distribution (in blue) of serum ferritin of the pooled data between 200 and 400 blood units and approximated normal distribution (in red).

### 3.3.3. Original dosing adjustment strategy

---

The dosing strategy proposed in the protocol of the DEEP-2 study consisted of the following steps:

(i) Assigning a starting dose to each patient. For deferiprone arm: patients will receive a starting dose of 75 mg/kg/day or their ongoing dosage as long as this does not exceed 100 mg/kg/day. For deferasirox arm: patients will receive a different starting dose of deferasirox depending on their current therapy:

- Naïve patients: 20 mg/kg/day
- Deferiprone treated patients: 20 mg/kg/day
- Deferoxamine treated patients: half deferoxamine dose (but not less than 20 mg/kg/day)
- Deferasirox treated patients: current patient's posology (but not higher than 40 mg/kg/day)

(ii) If there is an increase of more than 20% in serum ferritin compared to the previous visit at the central laboratory, increase the dose of a step of 12.5 mg/kg/day for deferiprone treated patients (until a maximum of 100 mg/kg/day) and of 5 mg/kg/day for deferasirox treated patients (until a maximum of 40 mg/kg/day).

### **3.3.4. Model-based dose adaptation strategy relying on long-term predictions**

The proposed model-based dosing strategy consisted of the following steps (Figure 3.7):

(i) A starting dose is assigned to each patient. Since the model-based strategy was tested on a cohort of virtual patients, their previous clinical history was unknown. Therefore, to consider a mixture of naïve patients and patients already on chelation therapy, three different approaches have been followed to assign the starting dose, which are:

- To all the patients the mean between 20 and 40 mg/kg/day for deferasirox and between 75 and 100 mg/kg/day for deferiprone was used as starting dose;
- A random value extracted from a uniform distribution between 20 and 40 mg/kg/day for deferasirox and between 75 and 100 mg/kg/day for deferiprone was used as starting dose;
- To half of the patients the mean between 20 and 40 mg/kg/day for deferasirox and between 75 and 100 mg/kg/day for deferiprone was used, while to the other half of the patients 20 mg/kg/day for deferasirox and 75 mg/kg/day for deferiprone was used as starting dose.

(ii) Empirical Bayes ("post-hoc") estimates (EBEs) of individual random effects are obtained in NONMEM by the maximum a posteriori (MAP) method using patient's monthly observations (collected during the monthly visits at the local laboratory) until the first visit at the central laboratory and

---

the population PK-PD model. Therefore, population parameters are fixed to their actual value (i.e. the one estimated from historical data and used in simulation). Then, the individualized model is used to extrapolate serum ferritin at 12 months for each patient.

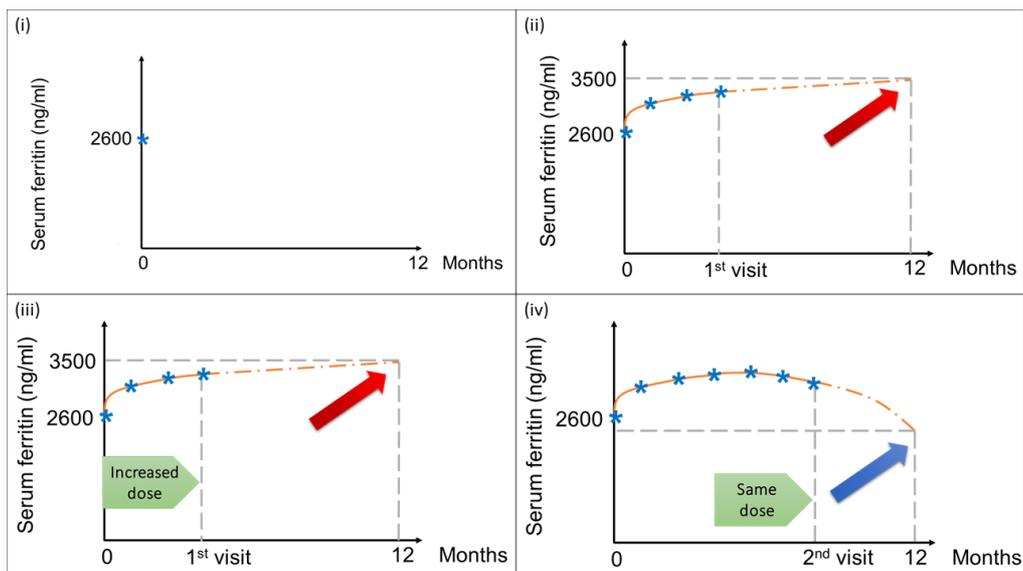
(iii) Using the serum ferritin baseline of each patient and the extrapolated value at 12 months, the rules defined in the protocol are applied to determine if the patient will be a responder or not. According to the protocol, the chelation therapy (with deferiprone or deferasirox) is considered successful, and, therefore, the patient is a responder, when serum ferritin after one-year treatment, compared to baseline, is decreased by 20% or more if baseline serum ferritin is  $\geq 2500$  ng/ml, or is decreased by any percentage or is increased less than 15% and serum ferritin within 12 months stay below 2500 ng/ml if baseline serum ferritin is  $< 2500$  ng/ml.

Then, if the patient is predicted to be a responder, he continues with his current dose; otherwise, the dose is increased of a step of 12.5 mg/kg/day for deferiprone, and of 5 mg/kg for deferasirox (as stated in the original DEEP-2 study protocol). He will then continue with this new dose until the next visit at the central laboratory.

(iv) Using the ‘true’ individual parameters of each patient and its new dosage, serum ferritin data of each patient are simulated from the first visit at the central laboratory to 12 months.

Repeat the steps from (ii) to (iv) before each visit at the central laboratory until the end of the trial (i.e. 12 months). Then, serum ferritin baseline and the serum ferritin value at the end of the trial are used to calculate the proportions of successfully chelated patients in the two treated arms.

Dose adjustment strategies were compared in terms of percentage of non-responders at the end of the trial. For each strategy, three different intervals between one visit at the central laboratory and the next, that are 1, 3, 4, and 6 months. For each strategy and for each visit-to-visit interval tested, 50 replicates were simulated. The Wilcoxon rank sum test was used to test the difference between the proportions of non-responders for the two strategies.



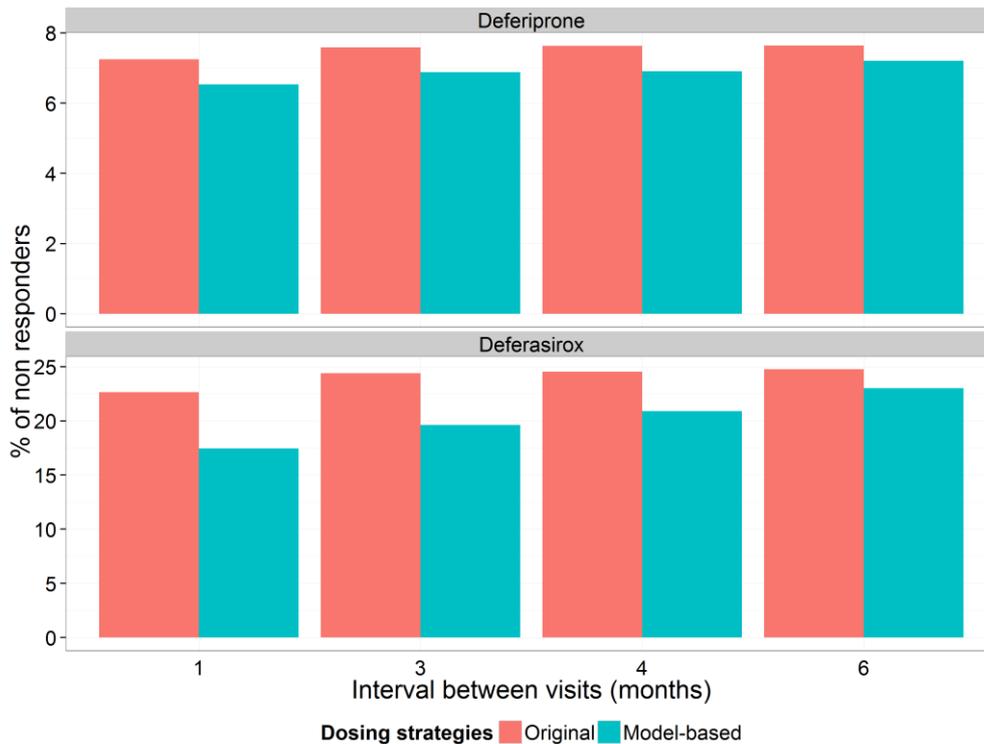
**Figure 3.7:** Steps (i)-(iv) for the model-based dose adaptation strategy relying on long-term predictions. Blue asterisks between one visit and the next represent monthly ferritin assessments at local laboratories. Dose can be adjusted only during the visits at the central laboratory (denoted in the picture as 1<sup>st</sup> visit and 2<sup>nd</sup> visit).

### 3.3.5. Results

The model-based strategy improved therapeutic efficacy in both the treated arms. In this case, the benefit was rather small, but still statistically significant, except for the case in which the monitoring was done only every six months (Table 3.3 and Figure 3.8). This analysis also showed that the results of the original strategy do not depend on the interval between one visit at the central laboratory and another, even if a close monitoring is preferred (Table 3.3 and Figure 3.8); three months can be considered a right balance between efficiency of the dosing adjustment strategy and feasibility. Sensitivity analysis concerning the starting dose has been performed. The impact of different starting doses on the percentage of success was investigated: all the three approaches adopted (see paragraph 3.3.4) did not affect the resulting outcome. Since in the original protocol, the total dose given to each patient was rounded, the impact of the type of rounding was evaluated. For deferiprone, the total daily dose is given three times daily and according to its corresponding value in ml, is rounded to a multiple of 2.5 ml or 0.2 ml if one-third of the total daily dose in ml is higher or lower than 10 ml, respectively (as stated in the protocol). For deferasirox, the total daily dose in mg is given once daily and is rounded to the nearest multiple of 125 mg, which is the smallest tablet size available in the trial (as stated in the protocol). Results in terms of percentage of successful runs were not statistically different for the no-rounding and the rounding scenarios (results not shown here).

**Table 3.3:** Results of the two dose adjustment strategies. L1: deferiprone; XJ: deferasirox.

Drug	Visit-to-visit interval (months)	Proportions of non-responders (%)		p-value
		Original strategy	Model-based strategy	
L1	1	7.25	6.54	<0.01
	3	7.59	6.89	<0.05
	4	7.63	6.91	<0.05
	6	7.64	7.21	>0.05
XJ	1	22.64	17.45	<0.001
	3	24.40	19.62	<0.001
	4	24.55	20.9	<0.001
	6	24.75	23.02	<0.01



**Figure 3.8:** Histograms of median of the proportions of success in the two arms for the original (red) and model-based (blue) dosing strategies and for different time intervals between one visit at the central laboratory and the next (i.e. 1, 3, 4, and 6 months).

### 3.4. Discussion

---

To explore the complexity of iron homeostasis and ferritin response following chelation therapies, the use of a quantitative approach for describing the dynamics of iron overload and its progression should be encouraged. In this Chapter, a new PK-PD model for iron overload has been developed. The dynamics of serum ferritin concentration in patients undergoing chelation therapy with deferiprone or deferasirox was successfully captured. This analysis highlighted the impact of disease- and patient-specific factors, such as the annual blood consumption and the serum ferritin baseline at the start of the treatment, on the trend of serum ferritin following chelation therapy.

CTS can be undergone with this newly developed PK-PD model to answer several clinical questions which are not yet fully explored, providing an invaluable tool in decision-making.

The first application of this PK-PD model consisted in investigating the possible benefits given by a model-based dose adjustment strategy, using as final endpoint the proportions of non-responders. In particular, the dose adaptation strategy proposed for the DEEP-2 non-inferiority study was compared to a novel model-based approach. CTS were performed to compare the two methods and it has been demonstrated that the model-based strategy improved therapeutic efficacy in both the treated arms. In this case, the benefit was rather small but still statistically significant. Besides, an interval of three months between one visit at the central laboratory and another seems to be adequate for the model-based approach, while with the original dosing adjustment strategies also six months can be considered because the percentage of non-responders is almost constant with 3, 4, and 6 months interval between visits at the central laboratory.

Limitations of the presented analysis include the fact that treatment compliance was not included since no quantitative data on that were available.

---

# Supplementary material to

## Chapter 3

---

### List of assumptions for dataset creation

George et al., 1994 (71): Study involving 24 patients with transfusion-dependent thalassemia from the Thalassaemic Clinic, Universiti Kebangsaan Malaysia, and from the Paediatric Department, General Hospital, Kuala Lumpur, from January 1989 to January 1992. The study population consisted in 18 Chinese and 6 Malay patients, aged from 2 to 11 years. A data profile chart was reported with patient ID, age (in years), sex, race, serum ferritin ( $\mu\text{g/L}$ ), and the number of blood transfusion units ( $\text{IU}=350 \text{ ml}$ ). Two patients were excluded because they were on chelation therapy with deferoxamine at a dose of  $35 \text{ mg/kg}$  3-5 times a week when the patients were 8 years old (B11HSF and B16LWK). To derive patient body weight a spline interpolation was performed based on patient's age, using a dataset reporting age and corresponding weight for Chinese population (94). To calculate the blood consumption in  $\text{ml/kg/year}$  the following formula was applied:

$$\frac{((n^\circ \text{ of units of blood in 1992}) - (n^\circ \text{ of units of blood in 1989}) \cdot (350 \text{ ml}))}{((\text{mean between weight in 1989 and in 1992}) \cdot (3 \text{ years}))}$$

Choundry et al., 2004 (74): Study involving 75 thalassaemic patients aged from 4 to 14 years, not on regular chelation therapy. Then, 30 patients received deferiprone at  $50 \text{ mg/kg}$ , 21 patients at  $75 \text{ mg/kg}$  daily, 24 patients were followed as controls. No weight and blood consumption were reported. Mean serum ferritin for the three arms at 0, 4, 8, and 12 months were reported. A mean weight of  $40.46 \text{ kg}$  was assumed, calculated as the mean weight of male and female children (from 4 years) and teenagers reported in (37). Since not specified, it was assumed that these patients received  $300 \text{ ml}$  of RBC (recommended amount of blood for children above 15 kg (95)) every 3 weeks. Therefore, the blood consumption per year was calculated as:  $((52/3) \cdot 300 \text{ ml})/(40.46 \text{ kg})$ .

Cappellini et al., 2006 (77): Comparative phase III study to demonstrate the efficacy of deferasirox in regularly transfused patients with beta-thalassemia aged 2 years or older. 296 patients were randomized to receive deferasirox and 290 to receive deferoxamine (arm not considered). The mean

---

age reported for the deferasirox arm was 17 years; a mean weight of 60 kg was assumed according to (37). The average daily doses for the one-year period was dependent on the baseline LIC: 6.2 mg/kg/day for LIC 3 mg Fe/g dw or less (n=15), 10.2 mg/kg/day for LIC between 3 mg Fe/g dw and 7 mg Fe/d dw (n=78), 19.4 mg/kg/day for LIC between 7 mg Fe/g dw and 14 mg Fe/g dw (n=84), and 28.2 mg/kg/day for LIC above 14 mg Fe/g dw (n=119). The only information reported about the blood consumption was that all patients received at least 8 blood transfusions per year; hence, they were assumed to receive from 2 to 4 units per month (7-14 ml/kg/month). The baseline serum ferritin in each group was derived from the baseline LIC of each group, assuming a linear relationship between ferritin and LIC with a proportionality coefficient equal to the ratio between reported median serum ferritin (2212 µg/L) and median LIC (11.3 mg Fe/g dw).

Cappellini et al., 2010 (78): Prospective, one-year, multicenter, open-label phase IIIb trial, involving 1744 patients with thalassemia (n=1115), myelodysplastic syndromes (n=341), aplastic anemia (n=116), sickle cell disease (n=80), rare anemias (n=43), and other transfused anemias (n=49). Since the mean age was 30.6 years, a mean weight of 70 kg was assumed (mean weight was not reported in the original publication). Median change in serum ferritin measured at 0, 2, 6, 9, and 12 months were reported for three groups of patients receiving <20 mg/kg/day (n=586), ≥20-<30 mg/kg/day (n=972), and ≥ 30 mg/kg/day (n=149), respectively. The corresponding median serum ferritin baselines for the three groups were 2608, 3165, and 5048 ng/ml, respectively. The blood consumption per year in each group was derived from the iron intake of each group, assuming a linear relationship between blood consumption and iron intake with a proportionality coefficient equal to the ratio between mean blood consumption per year (116.3 ml RBC/kg/year) and the corresponding iron intake (0.41 mg/kg/day) reported. Since the dose was changing with time, the mean actual deferasirox dose reported at 0, 3, 6, 9, and 12 months for each dosing group was considered.

Galanello et al., 2006 (79): Open-label, non-comparative phase II trial in pediatric patients with transfusion dependent beta-thalassemia major carried out in three centers in Italy and one center in France. Forty male and female patients aged 2-17 years were stratified in two groups: children aged 2 to <12 years (n=20) and adolescents aged 12-17 years (n=20), both treated with deferasirox for 48 weeks. The mean deferasirox dose was 11.3 mg/kg/day. A mean body weight of 24 kg was assumed for the children group (mean age 6.7 years), while a mean weight of 54 kg was chosen for the adolescents group (mean age 14.1 years). The blood consumption per year was derived as the mean number of transfusions in each group, assuming each transfusion being of 300 ml of RBC according to (95).

Piga et al., 2006 (80): Phase II study to test the tolerability and efficacy of deferasirox to those of deferoxamine in 71 beta-thalassemia with

---

transfusional hemosiderosis from four centers in Italy. Patients were randomized to take once-daily deferiasirox (10 or 20 mg/kg; n=24 in both groups) or deferoxamine (40 mg/kg, 5 days/week; n=23) for 48 weeks (arm not considered). The mean body weight of the 10 mg/kg-arm was 52.4 kg, while was of 50.7 kg for the 20 mg/kg-arm. The blood consumption was derived from the mean volume of blood given daily (0.34 ml RBC/kg/day), multiplying it by 365 days. Mean serum ferritin data from baseline to the end of the trial were reported for the two arms.

Hoffbrand et al., 1998 (75): Study comprising 38 patients with beta-thalassemia, 1 with Hb E/beta-thalassemia, 1 with sickle cell/beta-thalassemia, 4 with sickle cell anemia, 3 with sideroblastic anemia, and 4 with myelodysplastic syndrome, all treated with deferiprone at a dose of 75 mg/kg/day. Only serum ferritin data of 20 patients who continued deferiprone therapy for a mean of 39.4 months and were not affected by chronic liver disease caused by hepatitis C, were considered. Since the mean age was 33.27 years, a mean weight of 70 kg was assumed (mean weight was not reported in the original publication). The blood consumption was a weighted mean between the values reported: 150-200 ml RBC/kg/year for 36 patients, 250-300 ml RBC/kg/year for 2 patients, 260 ml RBC/kg/year for 1 patient, 120 ml RBC/kg/year for 1 patient, and 250-300 ml RBC/kg/year for 11 patients.

Aydinok et al., 2007 (72): Study comprising 24 patients with thalassemia major randomized to receive deferiprone at a daily dose of 75 mg/kg alone (n=12) or in combination with deferoxamine (n=12) at a dose of 40-50 mg/kg twice weekly (arm not considered). The mean deferiprone dose was 78.2 mg/kg/day. The mean age reported for the first arm was 15.9 years; hence, a mean weight of 60 kg was assumed according to (37). The blood consumption reported was 156 ml RBC/kg/year. Mean serum ferritin data from baseline to the end of the trial were reported.

Pennell et al., 2006 (76): Open-label trial conducted in four centers in Italy and Greece, involving 61 patients with beta-thalassemia major, randomized in two arms: n=29 treated with deferiprone, n=32 treated with deferoxamine (arm not considered). Deferiprone was initiated at 75 mg/kg/day and increased to the target of 100 mg/kg/day. Mean daily dose of deferiprone at 0, 4, and 8 months were taken from (96). Since the mean age was 25.1 years, a mean weight of 70 kg was assumed (mean weight was not reported in the original publication). The mean blood consumption reported was 152 ml RBC/kg/year. Mean serum ferritin data over time were taken from (96).

Ceci et al., 2002 (73): Study comprising 532 patients with thalassemia major (except 1 with thalassemia intermedia), treated with deferiprone at a dose of 75 mg/kg/day. Only data relative to the 151 subjects who completed the 3 years of treatment were reported. Since the mean age was 21.3 years, a

---

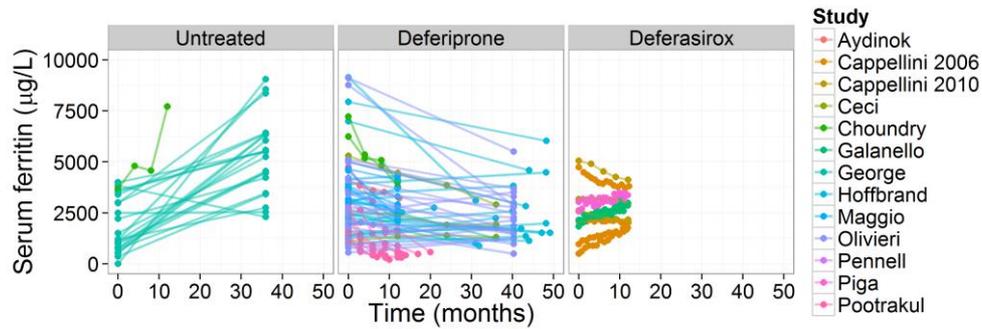
mean weight of 70 kg was assumed (mean weight was not reported in the original publication). The mean blood consumption reported was 151 ml RBC/kg/year. Mean serum ferritin data for three classes of basal ferritin levels, which are ferr<2000 (n=60), 2000<ferr<4000 (n=65), ferr>4000 (n=26), were reported.

Pootrakul et al., 2003 (83): Study involving 7 patients with Hb E/beta-thalassemia, and 2 with beta-thalassemia, treated with deferiprone at a dose of 50 mg/kg/day (n=7), 25 mg/kg (n=1), or from 25 to 50 mg/kg/day in 24 weeks (n=1). Individual dose reported for each of the seven patients. The patient #8 (with the dose scaled from 25 to 50 mg/kg/day) died before the end of the trial and was not considered in this analysis. Individual age and body weight were reported. The individual blood consumption was derived from the reported units of RBC taken during the study period, the individual body weight and the individual duration (weeks) in the study, assuming each packed unit of 300 ml. Individual serum ferritin profiles were reported.

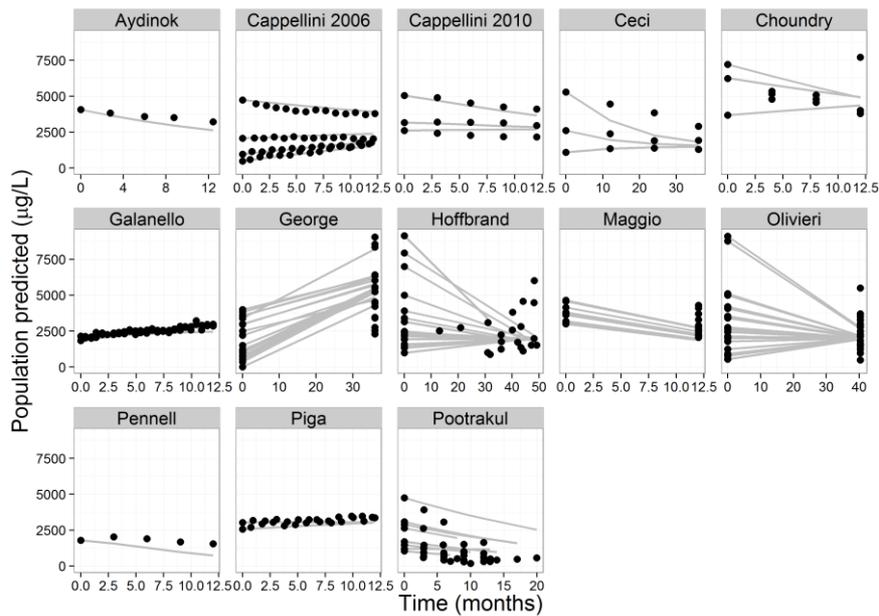
Maggio et al., 2002 (81): Individual serum ferritin concentration at baseline and at the end of the treatment (after 1 year) for 18 patients with baseline values >3000 ng/ml were reported. Only data relative to 11 patients treated with deferiprone (with a daily dose of 75 mg/kg) were considered. Since mean age was 20 years old, a mean weight of 64.15 kg (mean between 70.3 and 58 kg for male and female of 20 years old according to (37)) was assumed. The blood consumption was derived from total blood transfused during the study (ml), multiplying it by 60% (hematocrit value considered) to transform ml of blood in ml of RBC, and then dividing it by the mean body weight.

Olivieri et al., 1995 (84): Study involving 21 patients (with a mean age of 22 years) receiving deferiprone (75 mg/kg/day) for a mean of 3.1 years. A mean weight of 64.15 kg (mean between 70.3 and 58 kg for male and female of 20 years old according to (37)) was assumed. Each patient received transfusions to maintain Hb concentration above 10 g per deciliter, which corresponds approximately to 10 g of transfused iron yearly in a 70-kg adult. Since it is known that 200 mg of iron are present in a single unit of 285 ml of transfused RBC (61,77), the yearly blood consumption per body weight can be derived.

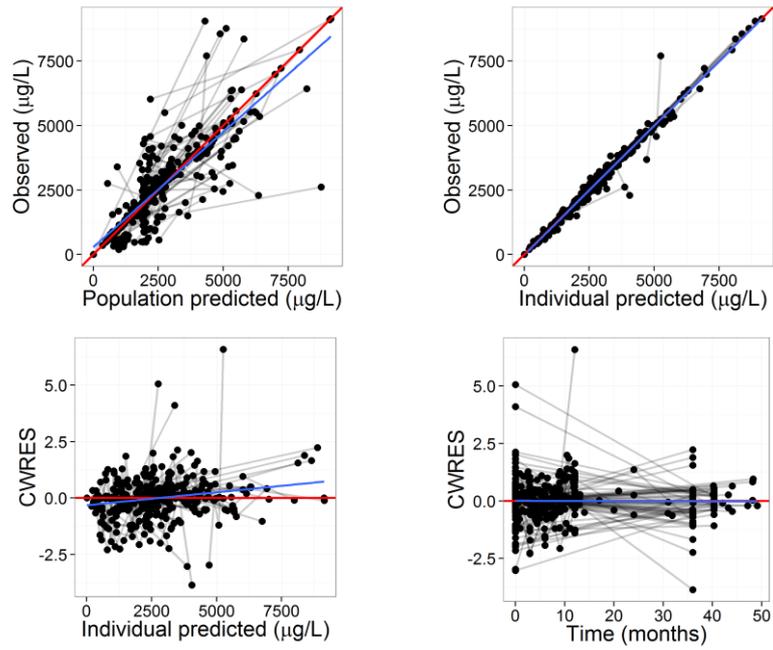
## Additional figures and tables



**Figure 3.S1:** Mean or individual serum ferritin concentration data collected from efficacy studies published in the literature, having arms treated with deferiprone or deferasirox or control arms.



**Figure 3.S2:** Plots of observed serum ferritin concentrations (black circles) and population predicted (grey line) for the final pharmacokinetic-pharmacodynamic (PK-PD) model.



**Figure 3.S3:** Goodness-of-fit (GOF) plots for the final pharmacokinetic-pharmacodynamic (PK-PD) model. Plot of observed serum ferritin concentrations vs. individual (top-right panel) and population predicted (top-left panel). Line of identity in red and regression lines of data points in blue. Plot of conditional weighted residuals (CWRES) vs. time (bottom-right panel) and individual predicted (bottom-left panel). Zero line in red and regression lines of data points in blue.

---

# Chapter 4

---

## Model-based optimization of pediatric efficacy trial duration and sample size

The trials in rare disease should be carefully designed to balance an appropriate trial duration and with an adequate patient participation to provide sufficient evidence for informing clinical decisions and give insights on the benefits and safety of treatments (16,97). When diseases are rare, patient recruitment and patient management imposes large efforts, especially in case of pediatric patients.

In this regard, M&S, in particular CTS, can be used to explore hypothetical “real-life” scenarios, ranging different experimental designs features, such as the population size and the trial duration. In this way, it is possible to predict ‘trial performance’ and identify possible limitations in the protocol design before enrolling children into the clinical trial (18,28,29).

In *Chapter 3*, the development and a possible clinical application of a PK-PD model for iron overload have been described. This model provided a reasonable basis for a more quantitative evaluation of the therapeutic intervention in the DEEP-2 study population.

In the first part of this Chapter, the possibility of an earlier prediction of clinical response was investigated via CTS with the aim of reducing the proposed original trial duration of 12 months. The choice of this specific trial length was driven by the fact that 12 months was arbitrarily considered as a sufficiently long time to see a stabilization in serum ferritin levels. Nonetheless, other relevant clinical questions have been addressed, e.g. how much time is required to observe an actual response.

---

Therefore, in the first part of this Chapter we evaluated to what extent prior knowledge together with PK-PD models can be used to drive predictions of response at long-term, allowing, therefore, shorter trial durations. Differently, in the second part of this Chapter, we explored the possibility of reducing the sample size, maintaining the original trial duration of 12 months.

## **4.1. Model-based methodology to reduce trial duration**

### **4.1.1. Simulation of a virtual pediatric population**

The population covariates were simulated according to the inclusion criteria specified in the protocol as described in *Chapter 3* (paragraph 3.3.1). Each patient was assigned to a dose randomly extracted from a uniform distribution between 75 and 100 mg/kg/day for deferiprone, and between 20 and 40 mg/kg/day for deferasirox. For the sake of simplicity, the same dose was maintained during all the trial duration.

### **4.1.2. DEEP-2 non-inferiority study simulation: shorter trial durations**

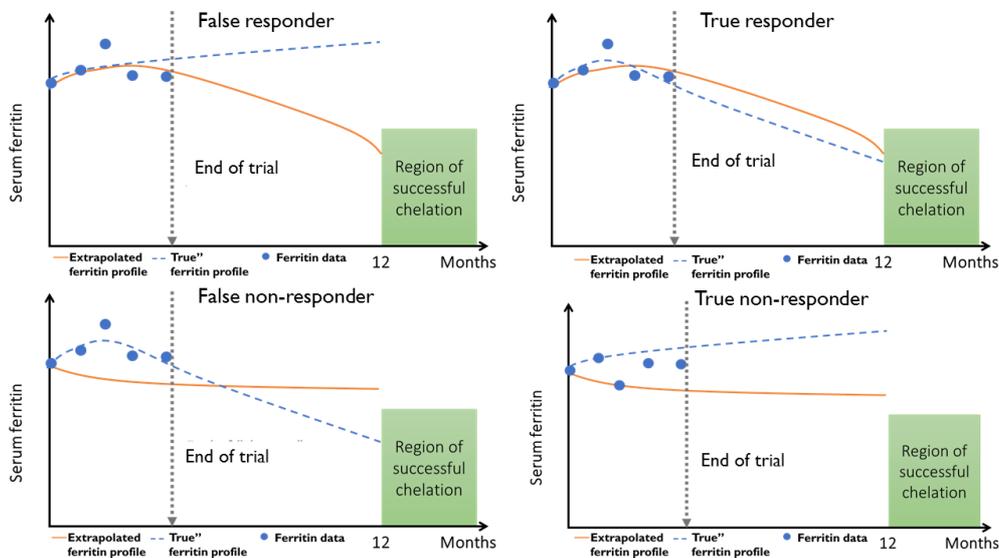
The possibility of an earlier prediction of clinical response was investigated via CTS. The changes to the PK-PD model initially estimated from literature data, which have been already described in *Chapter 3* (paragraph 3.3.2), were adopted also here.

The following procedure was carried on, consisting of the following steps:

(i) The PK-PD model is used to simulate serum ferritin data every month from 0 to 12 months for 388 virtual pediatric patients meeting the inclusion criteria of the protocol.

For each tested trial duration (from 1 to 11 months with a step of 1 month):

(ii) EBEs of individual random effects are obtained in NONMEM by the MAP method using patient's monthly observations (collected during the monthly visits at the local laboratory) from the start to the end of the trial and the population PK-PD model. Then, the individualized model is used to extrapolate serum ferritin at 12 months for each patient. Using the extrapolated value at 12 months and baseline serum ferritin levels, each patient is classified, following the criteria of success defined in the protocol, as true responder (true-positive, TP), true non-responder (true-negative, TN), false responder (false-positive, FP), or false non-responder (false-negative, FN) (Figure 4.1).

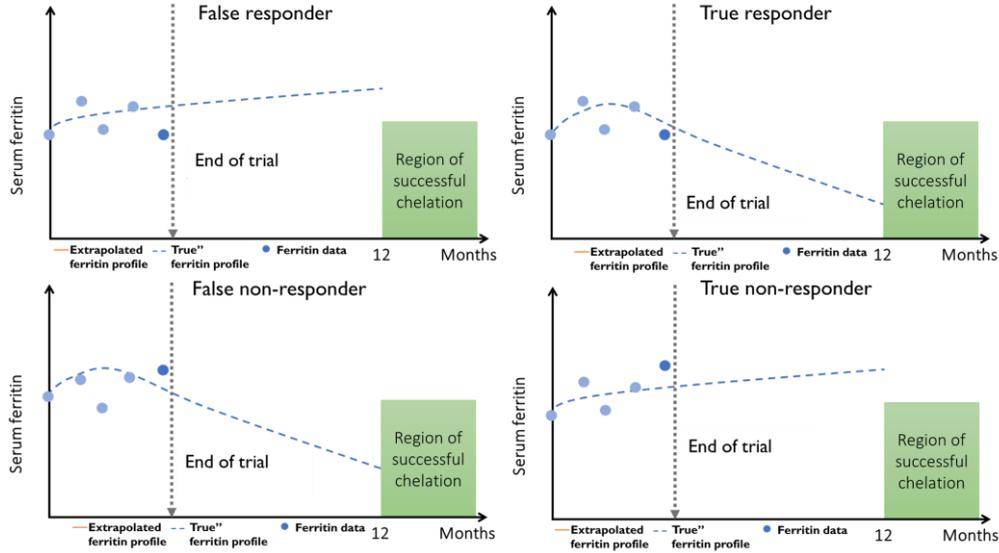


**Figure 4.1:** True (blue dashed line) and extrapolated (orange solid line) serum ferritin dynamics, and monthly measured serum ferritin data (blue circles) for a true responder, a false responder, a true non-responder, and a false non-responder.

The criteria reported in the original protocol stated: the chelation therapy (with deferiprone or deferasirox) is considered successful, and, therefore, the patient is classified as a responder, when serum ferritin after one-year treatment, compared to baseline, is decreased by 20% or more if baseline serum ferritin is  $\geq 2500$  ng/ml, or is reduced by any percentage or is increased less than 15% and serum ferritin within 12 months stays below 2500 ng/ml if baseline serum ferritin is  $< 2500$  ng/ml.

(iii) Using a non-model-based approach by comparing the baseline serum ferritin level and its actual value at the end of the trial, each patient is classified, following the criteria of success defined in the protocol, as a true responder (TP), true non-responder (TN), false responder (FP), or false non-responder (FN) (Figure 4.2).

The points (i-iii) have been repeated 200 times for each tested trial duration.



**Figure 4.2:** True (blue dashed line) serum ferritin dynamics, and monthly measured serum ferritin data (blue circles) for a true responder, a false responder, a true non-responder, and a false non-responder.

### 4.1.3. Comparing original and reduced trial durations

The objective of this analysis was to demonstrate the possibility, using a model-based approach, to detect in advance clinical response and, therefore, drive earlier conclusions on drug efficacy. To this aim, the results deriving from both the model-based approach and the non-model-based approach obtained for the different trial durations were compared in terms of sensitivity, specificity, positive-predictive-value (PPV), and negative-predictive-value (NPV), which are calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN}, \quad (4.1)$$

$$Specificity = \frac{TN}{TN + FP}, \quad (4.2)$$

$$PPV = \frac{TP}{TP + FP}, \quad (4.3)$$

$$NPV = \frac{TN}{TN + FN}. \quad (4.4)$$

They represent the capability of the model-based approach to discriminate between responders and non-responders based on the extrapolated value at 12 months. Sensitivity measures the proportions of responders that are correctly identified by the two strategies, while specificity measures the proportion of non-responders that are correctly identified. The PPV (or precision) is the ratio of true responders to combined true and false

---

responders and is more an information about the proportion of actual responder in the tested population than about the tested approach. Analogously, the NPV is the proportion of true non-responders in the tested population. The two approaches were also compared in terms of Type I and Type II errors for each tested trial duration. The Type I error represents the probability of false detecting an effect that is not present, while the Type II error is the probability of failure to recognize an effect that is present. Type I error is equivalent to the false positive rate (1-specificity), while Type II to the false negative rate (1-sensitivity). The following criteria for deciding when a certain trial duration was successful were adopted: 2.5<sup>th</sup> percentiles of sensitivity and specificity above 80%, 97.5<sup>th</sup> percentiles of Type I and Type II error below 20%, and 2.5<sup>th</sup> percentiles of NPV and PPV above 80%.

Besides, for the model-based approach, the ratios between the extrapolated value of serum ferritin at 12 months to its 'true' value were calculated for each patient in the simulated trials, for 200 simulated trials of each tested trial duration. The probability density curves of these ratios for each tested trial duration were then derived, and the proportions of each density area between 1/1.10 and 1.10 (10% variation from the 'true' value) were calculated. The smaller this area is, the more precise the extrapolation can be considered.

Two-way analysis of variance (ANOVA) was used to check if there were significant differences among the number of FP obtained with the different tested trial durations (only from 6 to 11 months), setting the significance limit ( $\alpha$ ) to 0.01. The choice of considering FP for this comparison was driven by the fact that they are the worst-case scenario in this analysis since they can lead to wrong conclusions about the non-inferiority between the two drugs. Then, if a significant difference was detected via two-way ANOVA, a post-hoc analysis was conducted using a Tukey's test ( $\alpha=0.01$ ) to compare all possible pairs of the tested trial durations.

#### **4.1.4. Results and discussion**

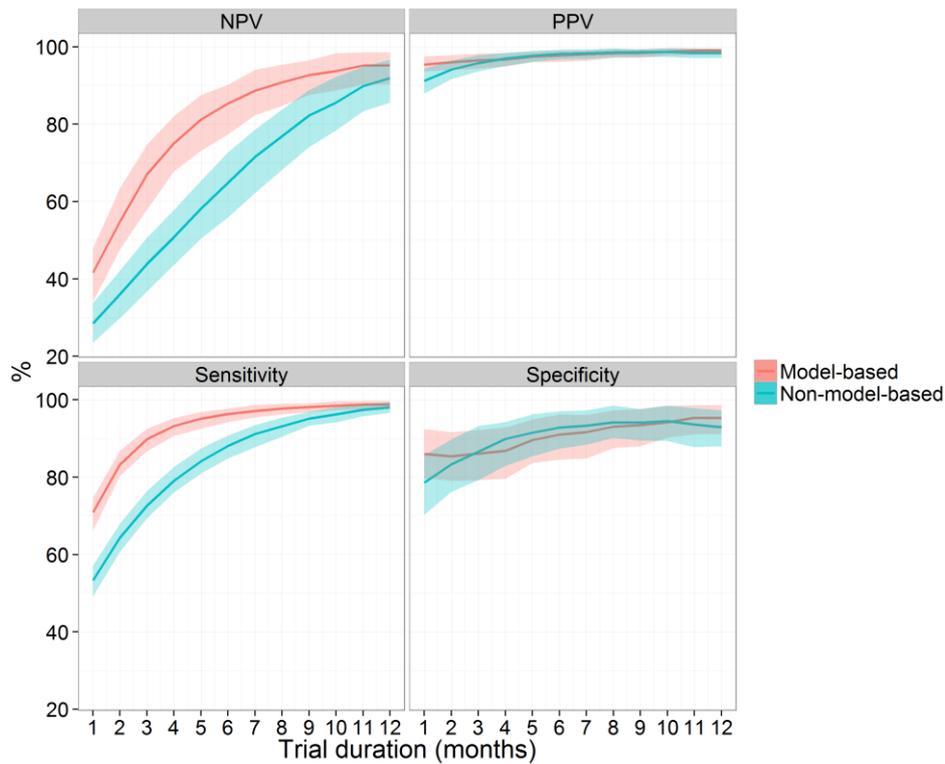
This work aimed, using a model-based approach, to detect in advance clinical response. Therefore, the PK-PD model presented in *Chapter 3* was used to extrapolate patients response using data collected in different trial durations. All these trial durations were compared in terms of sensitivity, specificity, NPV, and PPV (see criteria in paragraph 4.1.3), to derive how many months are required to observe a true response, and therefore to drive conclusions about drug efficacy.

The 2.5<sup>th</sup> percentiles of sensitivity, specificity, NPV, and PPV for the model-based approach at 6 months are all above 80% (Figure 4.3). This indirectly implies that, from the samples collected in the first 6 months, we are already able with high confidence to predict if a patient will be a non-responder at long-term (because of the high specificity). Always at 6 months, it is possible to detect with high confidence if a patient will be a responder at long-term (because of the high sensitivity) and, therefore, to draw

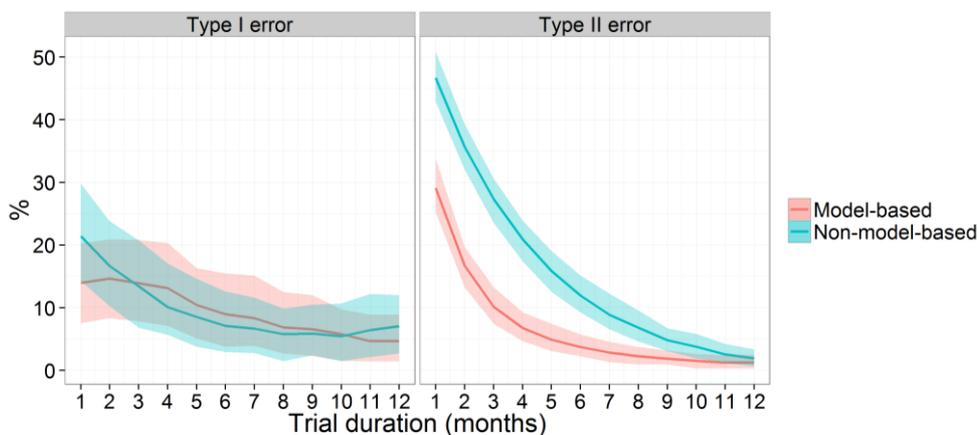
conclusions on the efficacy of the two drugs. Besides, the Type I and Type II errors in classifying patients between responders and non-responders is already acceptable at 6 months. In fact, there is only a slight decrease in Type I error after 6 months, and both the 97.5<sup>th</sup> percentiles of Type I and Type II errors stay below 20%.

Without a model-based approach, the 2.5<sup>th</sup> percentiles of sensitivity, specificity, NPV, and PPV stay above 80% only after 10 months. Besides, both the 97.5<sup>th</sup> percentiles of Type I and Type II errors after 10 months stay below 20%.

In conclusion, using a model-based approach is possible to predict at long-term if the chelation therapy will be successful (with small Type I and Type II errors) using only the first 6 months samples. The CTS with the non-model-based approach also suggests the possibility of reducing the original trial duration to 10 months, keeping the same strategy to evaluate treatment efficacy adopted in the original DEEP-2 study protocol.



**Figure 4.3:** Median (solid line) and 95% confidence interval (shaded area) of negative-predictive value (NPV), positive-predictive value (PPV), sensitivity, and specificity for the tested trial durations both for the model-based (in red) and non-model-based (in blue) approaches.



**Figure 4.4:** Median (solid line) and 95% confidence interval (shaded area) of Type I and Type II errors for the tested trial durations both for the model-based (in red) and non-model-based (in blue) approaches.

## 4.2. Model-based methodology to reduce sample size

In the previous section of this Chapter, we have investigated the potential reduction in trial duration, keeping the same sample size of the original DEEP-2 study. In the second part of this Chapter, we explored the possibility of reducing the sample size, maintaining the original trial duration of 12 months.

Challenges in studying rare disease in pediatric population include dealing with a limited availability of patients that can be enrolled, together with all the practical and ethical considerations that have to be made during pediatric patients management during the trial duration.

Consequently, the design of these studies necessitates the most informative analytical methods. To this aim, methods to assess appropriate sample sizes and incorporation of M&S in sample size calculations can be pursued to reduce the sample size and guarantee a Type II error below a certain threshold (98,99). For example, the StaR group has published six out of 11 planned standards where recommendations for practice are proposed (98). The fourth recommendation emphasized the need of a priori sample size determination during the study design.

Non-inferiority designs are adopted when the objective is to demonstrate that a certain treatment is “at least as good” or “not worse than” another competitor (100). Non-inferiority was the approach used for the design of the DEEP-2 study.

Here, using the DEEP-2 study as a driving example, we propose a sample size computational method, based on simulation of a population PK-PD

---

model, to find an appropriate sample size for a proper control of bias and, consequently, a meaningful interpretation of the results.

Besides, we show a sample size calculation method, based on a mixed model for repeated measures (MMRM), to see if repeated outcome measurements can increase study power compared to the original approach, also allowing a more precise estimation of variance between and within patients. MMRM methods are often applied when prior knowledge on the mechanism underlying drug response is missing, preventing the development of PK-PD models. However, MMRM methods do not incorporate any concentration-effect relationship, so they cannot be used to make inferences for age-related differences in PK, as in the case of pediatric population (85).

Finally, the model-based methods have been compared to the more classical approach where the outcome measure compared between the two groups is the proportion of successfully chelated patients at the end trial.

#### **4.2.1. Non-inferiority assessment method based on proportions of success (method 1)**

The DEEP-2 non-inferiority study was primarily conducted to demonstrate the non-inferiority of deferiprone to deferasirox in terms of percentage of successfully chelated patients after one-year treatment.

The analysis was based on a success criterion defined via a composite endpoint, including both serum ferritin and assessment of cardiac iron load by magnetic resonance image (MRI) (criterion B), that was applied to a pediatric population affected by hereditary hemoglobinopathies, requiring chronic transfusions and chelation.

For patients with less than 10 years of age, since cardiac MRI T2\* is uninformative and burdensome, the outcome measure was only based on serum ferritin measure (criterion A). In the absence of a model for MRI T2\* dynamics in this population, in the following analysis criterion A has been applied also to patients with more than 10 years of age.

In details, the following criteria reported in the original protocol has been used: the chelation therapy (with deferiprone or deferasirox) is considered successful, and, therefore, the patient is classified as a responder, when serum ferritin after one-year treatment, compared to baseline, is decreased by 20% or more if baseline serum ferritin is  $\geq 2500$  ng/ml, or is reduced by any percentage or is increased less than 15% and serum ferritin within 12 months stay below 2500 ng/ml if baseline serum ferritin is  $< 2500$  ng/ml.

The non-inferiority test originally adopted is a one-sided test ( $\alpha=0.025$ ) that test if the treatment difference between the experimental (deferiprone) and the standard (deferasirox) drug is no smaller than  $-\delta_{lim}$ . The margin  $\delta_{lim}=12.5\%$  was specified in the protocol. The null-hypothesis (H0) is that the difference  $\delta = p_1 - p_2$ , being  $p_1$  and  $p_2$  the proportions of success in the  $n_1$  and  $n_2$  patients randomized to deferiprone (experimental) and deferasirox (standard) arms, respectively, is less than  $-\delta_{lim}$ . Based on asymptotic

---

normal approximation for the difference of two binomial probabilities, the formula for the test statistic is:

$$Z_{NI} = \frac{(p_1 - p_2 + \delta_{lim})}{\sqrt{Var(p_1 - p_2)}} \quad (4.5)$$

where:

$$Var(p_1 - p_2) = \frac{p_1 \cdot (1 - p_1)}{n_1} + \frac{p_2 \cdot (1 - p_2)}{n_2}. \quad (4.6)$$

#### 4.2.2. Power calculation (method 1)

The formula used to calculate the total sample size  $n = (n_1 + n_2)$  given Type I error ( $\alpha$ ), power ( $1-\beta$ ), true percent of success (%) in the experimental group ( $p_1$ ) and in the standard group ( $p_2$ ), and the non-inferiority limit ( $\delta_{lim}$ ), is:

$$n = f(\alpha, \beta) \cdot \frac{[p_2 \cdot (100 - p_2) + p_1 \cdot (100 - p_1)]}{(p_2 - p_1 - \delta_{lim})^2} \quad (4.7)$$

where:

$$f(\alpha, \beta) = [\Phi^{-1}(\alpha) + \Phi^{-1}(\beta)]^2 \quad (4.8)$$

with  $\Phi^{-1}$  the inverse of the cumulative distribution function (CDF) of a standardized normal distribution.

To compare method 1 with the following proposed model-based approaches, the actual proportions of success calculated on a large pool of 10,000 simulated individuals (that will be used in the other two methods) have been used for  $p_1$  and  $p_2$ , using the same distribution of covariates and the PK-PD model described in *Chapter 3*. The proportions of success were 90% and 70% for the deferiprone and the deferasirox arms, respectively.

Reversing the Eq. 4.7, it is possible to calculate the Type II error ( $\beta$ ), and therefore the power ( $1 - \beta$ ), given different sample size values. The following sample size values were considered in this analysis: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400. The formula for the type II error calculation is:

$$\beta = 1 - \Phi \left( \sqrt{\frac{n}{[p_2 \cdot (100 - p_2) + p_1 \cdot (100 - p_1)]}} - \Phi^{-1}(\alpha) \right) \quad (4.9)$$

where  $\Phi$  is the CDF of a standardized normal distribution.

---

### 4.2.3. Non-inferiority assessment method based on repeated measurement (method 2)

In the DEEP-2 non-inferiority study, serum ferritin trends over the one-year period of treatment have also been investigated using a MMRM. Only samples collected quarterly at the central laboratory (at 3, 6, 9, and 12 months) were considered for this analysis as stated in the protocol (Table 2.S1 in *Supplementary material to Chapter 2*), and the corresponding change from baseline derived for each time and subject. The use of this derived statistic (i.e. change from baseline) makes the difference between the two drugs clearer. In fact, if absolute values of ferritin are considered, and a difference between the two drugs is found, this could be due either to a genuine effect of that drug or to a regression to the mean (i.e. a sharp decrease is usually observed in the group which has higher starting values).

In a repeated measure analysis of variance, two types of effects are considered: a between-groups (or treatment) effect and a within-subjects (or time) effect. Also, the within-subject correlation across repeated measures and the interaction between treatment effect and the elapsed time have been considered.

The following linear mixed-effects model was therefore fitted (100):

$$Y_{ijt} = \mu + \gamma \cdot TR_t + \tau_j \cdot time_j + \theta_j \cdot TR_t \cdot time_j + \varepsilon_{ijt} \quad (4.10)$$

Where:

- $Y_{ijt}$  is serum ferritin change from baseline for the  $i$ -th subject,  $j$ -th time and treatment  $t$ ;
- $\mu$  is an “overall” serum ferritin change mean;
- $\gamma$  represents the treatment effect;
- $TR_t$  is a binary indicator equal to 1 for deferiprone and to 0 for deferasirox;
- $\tau_j$  ( $j = 3,6,9,12$  months) represents the repeated measure time effect;
- $time_j$  ( $j = 3,6,9,12$  months) is a binary indicator equal to 1 for the  $j$ -th time and 0 otherwise;
- $\theta_j$  ( $j = 3,6,9,12$  months) represents the treatment by time interaction;
- $\varepsilon_{ijt}$  is the error term (or residual), i.e. the difference between the model prediction and the observed data, for the  $i$ -th subject,  $j$ -th time and treatment  $t$ .

An alternative form of the model above, in which the intercept and the treatment variable are removed, is:

$$Y_{ijt} = \tau_j time_j + \theta_j TR_t time_j + \varepsilon_{ijt} \quad (j = 0,3,6,9,12 \text{ months}) \quad (4.11)$$

where the estimated  $\theta_j$  in this case directly estimate the difference between the two groups at each respective time point (100).

---

Measures on the same subjects are likely to be correlated, therefore, a covariance matrix is used to consider these correlations. Since the measures that are close together in time are likely to be more correlated than measures further apart, a first-order autoregressive covariance structure was adopted. Thus, the covariance between measures at time  $j$  and  $k$  in subject  $i$  assigned to treatment  $t$  is:

$$cov(Y_{ijt}, Y_{ikt}) = \sigma^2 \cdot \rho^{|j-k|}. \quad (4.12)$$

The hypothesis that deferiprone is non-inferior to deferasirox at any of the measured times is assessed testing whether the treatment effect is consistent over time, which implies evaluating first if the interaction between  $TR_t$  (treatment) and  $time_j$  (time) is equal to zero. The rejection of this null hypothesis means that the interaction term is statistically and clinically significant; hence, this will be supportive of different trends between deferiprone and deferasirox arms, and the treatment effect should be evaluated time by time and not overall.

#### 4.2.4. Power calculation (method 2)

The procedure implemented to calculate Type II error (and therefore power) given a certain sample size consisted in the following steps:

(i) For each sample size  $n$ , extract randomly from the pool of 10,000 simulated patients,  $n/2$  patients treated with deferiprone, and  $n/2$  treated with deferasirox with their corresponding serum ferritin data (at 3, 6, 9, and 12 months). The dose was assigned randomly to each patient, extracting it from a uniform distribution between 75 and 100 mg/kg for deferiprone, and between 20 and 40 mg/kg for deferasirox; the same dose was maintained for all the trial duration, which was assumed here to be the original one of 12 months.

(ii) Use the data extracted at (i) to fit the model described in Eq. 4.10.

(iii) Assess the treatment by time interaction by testing whether the vector of interaction terms  $\theta_j$  is equal to 0 through an F-test ( $\alpha=0.025$ ).

(iv) Obtain estimates of treatment effect at specific times using the alternative form of the model (Eq. 4.11), in which the intercept and the treatment variable are removed.

(v) Perform on each  $\theta_j$  a non-inferiority t-test. The null hypothesis ( $H_0$ ) is  $\theta_j > \delta_{lim}$  while the alternative hypothesis ( $H_1$ ) is  $\theta_j \leq \delta_{lim}$ , where  $\delta_{lim}$  defines the non-inferiority region  $[-\infty; \delta_{lim}]$ . The test statistics is calculated as:

$$T_{NI} = \frac{\theta_j - \delta_{lim}}{\widehat{SE}_{\theta_j^*}} \quad (4.13)$$

where  $\theta_j$  is the treatment by time effect at the  $j$ -th time estimated in (iv), and  $\widehat{SE}_{\theta_j}$  is the corresponding standard error.

The non-inferiority margin  $\delta_{lim}$  is the maximum accepted difference between change from baseline of deferiprone and deferasirox, meaning that the difference between the curves of the change from baseline vs. time of deferiprone and deferasirox can be at most  $\delta_{lim}$  in each time point considered. The margin of 400 ng/ml was specified in the protocol. The Holm-Bonferroni correction procedure for multiple comparisons was used (101). It consists in ordering the  $p$ -values of all the tests from the lowest to the highest, and compare them with the corresponding significance criteria, calculated as  $\alpha/k$ ,  $\alpha/(k-1)$ ,  $\alpha/(k-2)$ , ...,  $\alpha/(k-1)$ , with  $k=4$  (i.e. the number of comparisons) and  $\alpha=0.025$ . If each  $p$ -value is less than the corresponding criterion, the non-inferiority of deferiprone to deferasirox is concluded for each time point.

The procedure described in the points (i)-(v) has been repeated 1,000 times. For each iteration, the Type II error was derived as the number of incorrect retains of the null hypothesis divided by the number of total iterations (i.e. 1,000).

To derive the Type I error, the same procedure was repeated, but in this case, in (i), the  $n$  patients were extracted from the same arm, and the number of incorrect rejections of null hypothesis divided by the total number of iterations (i.e. 1,000) was the Type I error. The Type I error is expected to stay always below the chosen  $\alpha$  of 0.025, independently from the sample size.

#### **4.2.5. Non-inferiority assessment method based on model-based approach (method 3)**

A sample size computational method, based on simulation of a population PK-PD model, was used, and the Type II error (and therefore power) obtained for each tested sample size was calculated.

The procedure described hereafter was followed:

(i) For each sample size  $n$ , extract randomly from the pool of 10,000 simulated patients  $n/2$  patients treated with deferiprone, and  $n/2$  treated with deferasirox with their corresponding serum ferritin data (at 3, 6, 9, and 12 months). The dose was assigned randomly to each patient, extracting it from a uniform distribution between 75 and 100 mg/kg for deferiprone, and between 20 and 40 mg/kg for deferasirox; the same dose was maintained for all the trial duration, which was assumed here to be the original one of 12 months.

---

(ii) Use the data extracted at (i) to fit the PK-PD model described in *Chapter 3* (paragraph 3.3.2).

(iii) Compare the two CDF of the effect of deferiprone and deferasirox. Given the assumed model structure, the drug effect was described as a first-order elimination from the iron compartment (see *Chapter 3*, paragraph 3.2.3). Therefore, the distributions of the products between the patients individual slope parameters and steady-state drug average concentration have been compared. To this aim, a two-sample Kolmogorov-Smirnov test ( $\alpha=0.025$ ) was performed. The alternative hypothesis here is that the distribution of the effect of deferiprone is stochastically greater than the distribution of the effect of deferasirox (i.e. the CDF lies below and hence to the right of the deferasirox one).

The procedure described at the points (i)-(v) has been repeated 1,000 times. For each iteration, the Type II error was derived as the number of incorrect retains of the null hypothesis divided by the number of total iterations (i.e. 1,000).

To derive the Type I error, the same procedure was repeated, but in this case, at point (i), the  $n$  patients were extracted from the same arm, and the number of incorrect rejections of null hypothesis divided by the total number of iterations (i.e. 1,000) was the Type I error. The Type I error is expected to stay always below the chosen  $\alpha$  of 0.025, independently from the sample size.

#### **4.2.6. Results and discussion**

The DEEP-2 non-inferiority study aimed to compare the efficacy of deferiprone and deferasirox in terms of percentage of success in the two groups. The success of the chelation therapy was originally evaluated using only baseline serum ferritin and the last measured value at the end of the one-year treatment (method 1). Since serum ferritin trends were also recorded, MMRM analysis (method 2) was also investigated in this second section of Chapter 4. The availability of a PK-PD model for iron overload also allowed the possibility of testing a model-based approach (method 3) to find the required sample size to guarantee a certain power and significance limit.

The sample size obtained from this analysis cannot be compared with the sample size originally planned since this was derived assuming proportions of success different from the ones used here. In the original protocol, in fact, the success of the chelation therapy was based for those patients aged  $> 10$  years old on a composite endpoint, comprising both the serum ferritin decrease from baseline and the change in T2\* from baseline. Given the absence of a PK-PD model linking deferiprone or deferasirox concentration to T2\* dynamics, this composite endpoint was not considered here but only change from baseline at the end of the one-year trial. The results of this

---

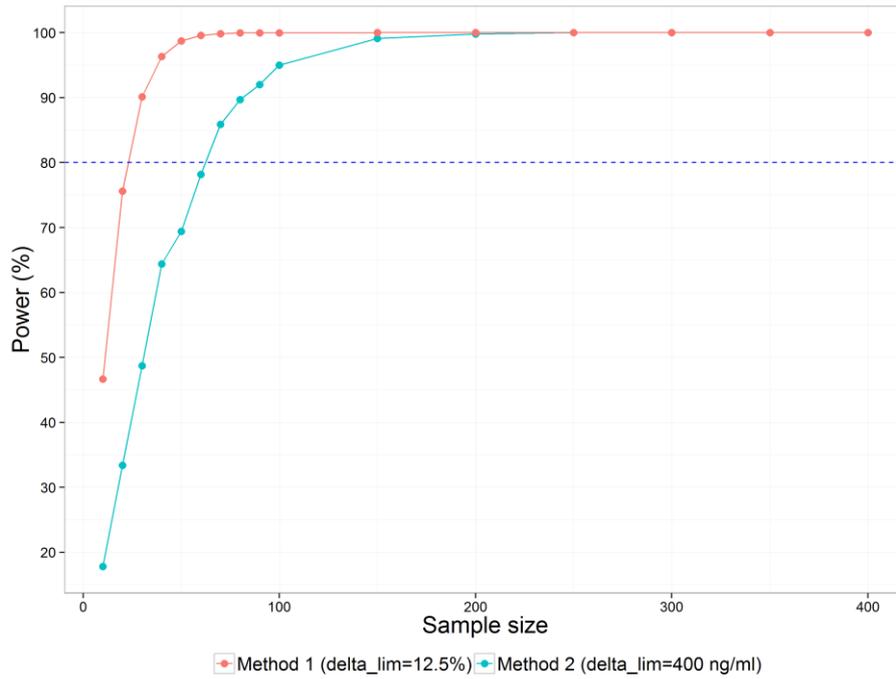
analysis are therefore intended to compare three different approaches, in terms of Type II and Type I errors, but they are not meant to provide any conclusions on the sample size of the ongoing DEEP-2 non-inferiority study.

Besides, method 3, in which the distributions of the effect in the two arms are compared through a two-sample Kolmogorov-Smirnov test, assesses if the distribution of the effect of deferiprone is not statistically smaller than the deferasirox one. This alternative hypothesis is different from the ones tested in method 1 and 2; in fact, in these cases, the alternative hypothesis is that the effect of deferiprone is at least not worse than the effect of deferasirox, given a certain threshold  $\delta_{lim}$ . To allow a comparison between method 3 and the other methods, a threshold limit of 0 was assumed, i.e. the null hypothesis tested is that deferiprone is better than deferasirox.

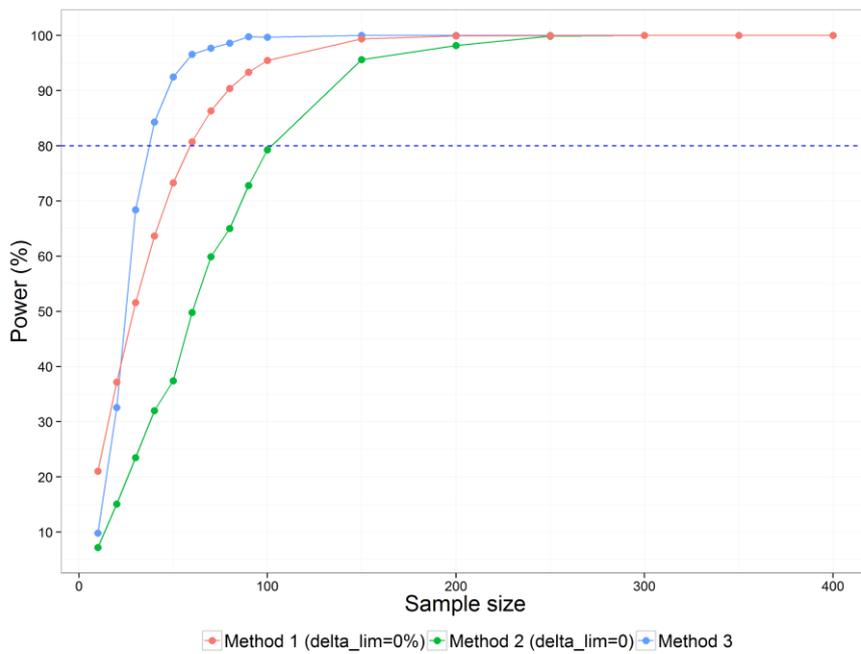
A comparison between method 1 and method 2 with their original threshold is reported in Figure 4.5. With method 2, a larger sample size is requested to guarantee 80% power, but the conclusion derived with this method is much stronger than the one obtained with method 1. In fact, with method 2 we tested the hypothesis that deferiprone is non-inferior to deferasirox during all the 12 months, from the very beginning of the treatment (i.e. 3 months) to the end, while with method 1 we can only conclude that after one-year the two treatments are comparable.

In Figure 4.6 a comparison between all the three methods is shown. method 3 guarantees the same power with a lower sample size, compared to method 1 and 2 where the non-inferiority margin  $\delta_{lim}$  was set to 0 to allow the same hypothesis testing.

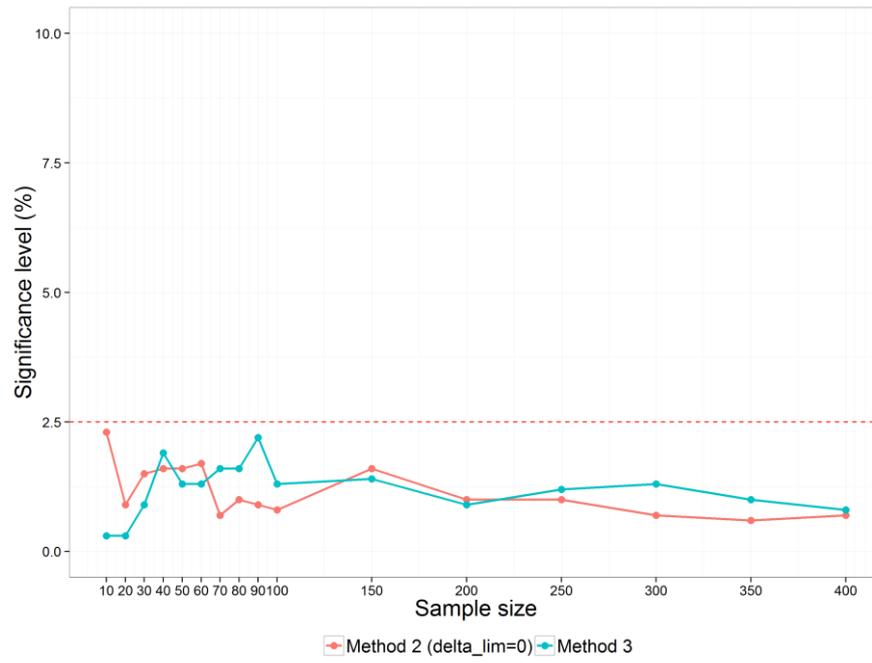
In Figure 4.7 a comparison between the significant limit or Type I error for method 2 and 3, both assessed via CTS, is shown. As expected, the significance level is constant for all the tested sample sizes, but it cannot be concluded that one method is better overall than the other.



**Figure 4.5:** Power (%) vs. sample size for method 1 (in red) and method 2 (in blue) with  $\delta_{lim}$  equal to 12.5% and to 400 ng/ml for method 1 and 2, respectively.



**Figure 4.6:** Power (%) vs. sample size for method 1 (in green), method 2 (in blue), and method 3 (in red) with  $\delta_{lim}$  equal to 0 for both method 1 and method 2.



**Figure 4.7:** Significance level (%) vs. sample size for method 2 with  $\delta_{lim}$  equal to 0 (in blue) and method 3 (in red).

---

# Chapter 5

---

## Overall Conclusions

Throughout this thesis we have focused on the importance of using prior knowledge, together with M&S, to improve evidence generation and evidence synthesis in case of pediatric trials in rare diseases. We have also demonstrated how long-term model-based predictions of efficacy derived from a PK-PD model, entirely developed from historical data, can be used to drive dosing adjustment decisions. Another important feature of the work presented in this thesis is the use of CTS to exploit benefits-risks of different designs and situations, before starting the real trial. The design of informative trials is particularly important for studies in children because ethical and practical constraints impose limits on sample size and sampling windows.

All these critical aspects have been investigated throughout the chapters of this thesis, using beta-thalassemia and iron chelation therapy to treat iron-overload complications deriving from life-long transfusions, as a paradigm for rare disease and associated drug therapy. Besides, the DEEP-2 study, comprising also a PK sub-study, was used as a case study throughout this thesis.

In *Chapter 2*, the importance of augmenting the sparse data collected in a pediatric study with prior information derived from similar studies on the same drug in adults have been demonstrated. Without the use of priors, the probabilities of being able to reach a successful minimization of the NLME estimation algorithm are only around 10%. The use of weakly-informative priors has proved to increase the robustness for model identifiability, allowing the integration of prior knowledge from historical data without dominating the estimation method, which is of great importance in situations where data collected are sparse. Nevertheless, even if weakly-informative priors are used, we are conscious that with only one sample per patient intra-individual variability can be hardly be distinguished from IIV.

---

Therefore, we have demonstrated how knowledge integration can be used to design pediatric trials to obtain quality and informative data, always considering the limitations due to feasibility reasons.

In *Chapter 2*, the focus was on the optimization of sampling times with the aim of finding the best sampling time windows starting from the prior knowledge available. Then, we highlighted the possible risks of drawing wrong conclusions on PK of an iron chelator in children, comparing the original protocol with several other new optimized protocols where, differently from the original one, more than one sample is collected for each patient. Therefore, we showed the important drawbacks of collecting only one sample in the original protocol since they lack the necessary informative content, leading to a probability of more than 60% of over/underestimating the exposure of more than 25%. Furthermore, we have demonstrated how the use of ED-optimization methods together with a slightly increased number of samples per subject (i.e. 3 samples) can lead to very accurate and precise predictions of drug exposure in the pediatric population, starting from prior knowledge available mainly in the adult population.

The integration of prior knowledge with oncoming data is important not only for evidence synthesis and data generation, but also to describe the progression of a disease and, therefore, to predict long-term patient clinical outcome.

To this aim, in *Chapter 3* a PK-PD model for iron overload based on literature data was developed. This model was used to investigate the possible advantages of a model-based dosing adjustment strategy compared to the original strategy where only the serum ferritin change from the previous visit at the central laboratory was used instead. The model-based approach improved therapeutic efficacy. Even if the benefit was rather small, it was however still statistically significant, except for the case in which the monitoring was done only every six months. The best time interval for monitoring the patients and adjusting their doses was found to be three months for the model-based strategy, while the original strategy seemed not to benefit from a closer monitoring in time.

Finally, in *Chapter 4* the PK-PD model was used to perform CTS to investigate the possibility, using model-based predictions, of reduced trial duration or sample size. We have demonstrated that, using a model-based approach, is possible to predict at long-term if the chelation therapy will be successful (with small Type I and Type II errors) based only on the samples collected in the first six months. Also with a non-model-based approach, we highlighted the possibility of reducing the original trial duration to ten months since serum ferritin dynamics seemed to be stabilized already at that time.

Besides, the model-based approach proved to guarantee the same power with a lower sample size, compared to original method used to assess non-inferiority in the DEEP-2 study and to the use of MMRM methods.

---

Starting from some the limitations associated with the study of drugs in a pediatric rare disease context, innovative methodologies to face these challenges, such as the integration of existing information, the optimization of study design, CTS, have been proposed and successfully applied to a real-life study. Such approaches can be extended or adapted for the study of other orphan drugs, informing people decisions throughout all their development process.

---

# Appendix **A**

---

## **Complex Bayesian modeling workflows encoding and execution made easy with a novel WinBUGS plugin of the DDMoRe IOF**

### **A.1. Introduction**

Bayesian modeling could be used to incorporate prior information derived from previous studies (e.g. studies in adults) to support data analysis and reducing uncertainty on model parameters (see *Chapter 2*), borrowing strength from historical data, especially in those cases where the sample size available is limited (e.g. pediatrics, rare disease, special populations) (102). Besides, prior information can be used for design optimization of new trials in children (103), both via ED-optimization techniques (see *Chapter 2*) or CTS (see *Chapter 4*).

Bayesian methods also allow the propagation of uncertainty through the different hierarchical levels of a model or among different models and enable direct probabilistic inferences on the posterior distributions (104,105).

Different software tools, such as WinBUGS (106,107), OpenBUGS (108,109), Stan (110,111), JAGS (112,113), and NONMEM (114,115), can be used to encode Bayesian models and to carry out parameter estimation via Markov Chain Monte Carlo (MCMC) algorithms (116). WinBUGS enables flexible statistical model specification and relies on additional tools, such as the WinBUGS Development Interface (WBDev) (117) with the BUGSModelLibrary (118,119), to cover many features required in pharmacometric modeling, such as custom ordinary differential equations (ODEs), IF-THEN-ELSE statements, definition of custom PK models and dosing schedules, not directly available in the BUGS language (108,120).

---

The described add-ons can be integrated within WinBUGS and enable the encoding of customized functions in Component Pascal language (121), including ODE specification, and support the use of NONMEM-formatted data items.

Considering the other popular or emerging modeling tools mentioned above, although enabling to run several model classes, the efficient implementation of PK-PD models with ODEs and dosing schedules is limited (or missing) in JAGS, Stan and OpenBUGS (122), even if Stan is currently in further development, and it seems to be a promising tool. The most recent versions of NONMEM, the most widely used software for population analysis via maximum likelihood approach, also enable Bayesian analysis via MCMC methods (Gibbs/Metropolis-Hastings and Hamiltonian/No U-Turn Sampling). Despite NONMEM has unique advantages for Bayesian analysis, e.g. parallel computation enabling within-chain parallelization, and more flexibility has also been given to users with the last release (v.7.4) in terms of prior distributions choice, WinBUGS is recommended when more than two levels of variability or an expanded choice of prior distributions are desired (120,122–125). For these reasons, the WinBUGS suite described above represents a key option for Bayesian modeling in the PKPD context (125–130).

It is worth noting that the WinBUGS suite enables the encoding of complex models, but a significant encoding effort is required, including model and functions definitions via BUGS and Component Pascal languages. Other packages, such as PKBugs (131,132), Pharmaco (117), and the BUGSModelLibrary (118,119) have been proposed to facilitate pharmacometric models encoding, but they are limited to a set of predefined compartmental models, and the development of more complex ones still requires significant encoding efforts as described above (125).

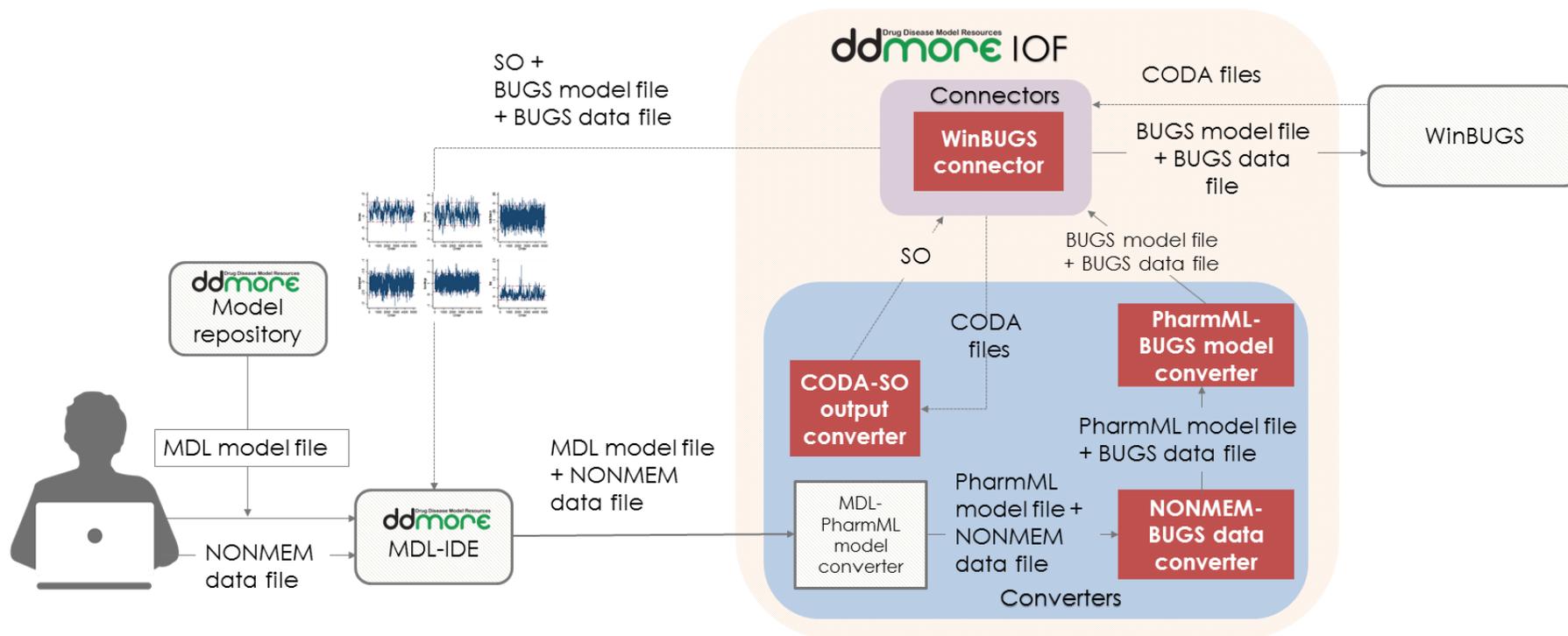
The DDMoRe IOF (Figure A.1) (133,134) is a software infrastructure developed by the DDMoRe consortium (135) and now supported by the DDMoRe Foundation (136), aimed to facilitate the exchange and integration of models across different languages or tools. The IOF has two key system-to-system target tool-independent interchange standards: PharmML, a XML-based computer language for model representation (137), and SO, a storage format for pharmacometric analysis results (138). The IOF can be accessed via a graphical user interface, the MDL Integrated Development Environment (MDL-IDE) (133), where the user can encode models in MDL (139), and script workflows in R programming language (140). MDL is a declarative human-readable/writeable language, characterized by a modular object-based structure, that is used to represent the information required to describe models (139). MDL facilitates model definition and, for use in Bayesian estimation, definition of prior distributions for parameters. Specific R functions, available in the *ddmore* R package [33], support model definition by composing different MDL objects and enabling the execution of the desired modeling tasks. A set of converters and connectors, described in Figure A.1, perform the MDL-to-PharmML and the PharmML-to-target

---

tool automatic translation, and the execution of a desired task, respectively. Finally, results are provided back to users via SO.

The standardized nature of languages, functions, and outputs in the IOF can significantly alleviate the burden of model/dataset encoding or recoding in different target languages for allowing the exploitation of the different features made available from the different software tools (137). It can also support the reproducibility of results and the interoperability among modeling tools, which are long-standing problems in pharmacometrics, to eventually streamline complex workflows (137,141,142).

In this Appendix, we aim to present a novel WinBUGS plugin for the IOF (134) and demonstrate its usefulness in the programming and execution of a previously published diabetes-linked Bayesian modeling workflow. This work has been recently submitted to *Computer Methods and Programs in Biomedicine*. The developed software framework will provide a solution to interoperability issues in Bayesian modeling and to the currently difficult encoding of complex PKPD models in WinBUGS. The IOF now supports a wide range of tools for estimation (Monolix, NONMEM, WinBUGS), diagnostics (PsN, Xpose), simulation (Simulx, SimCyp) and optimization (PopED, PFIM) (134).



**Figure A.1:** Information flow of the DDMoRe interoperability framework. See full caption in next page.

**Figure A.1:** Information flow of the DDMoRe interoperability framework. The DDMoRe interoperability framework (IOF) is an integrated set of converters and connectors for many common programming tools and languages. Together with the IOF, the translation of models to different software tools is provided by the integration of two standard languages: Model Description Language (MDL) and Pharmacometric Markup Language (PharmML). A user-interface, called MDL-Integrated Development Environment (MDL-IDE), allows the user to create and edit files containing MDL code. Alternatively, the user can retrieve and use PharmML and MDL model codes of a variety of state-of-the-art models in key therapeutic areas freely and publicly available in the DDMoRe Model Repository (143). Once the MDL model code is available, the user can run a specific task (estimation/simulation) in one of the programming tools integrated in the IOF (e.g. WinBUGS) via R code, also specifying the settings which will be passed on to the target tool (variables to be monitored, number of chains, number of updates in the Markov Chain, etc.). Then, three automatic translations are performed in the background: (i) MDL to PharmML model translation, (ii) NONMEM-formatted to BUGS data file translation, (iii) PharmML to WinBUGS model translation, which generates all the necessary model files, including BUGS and Component Pascal files. Then, a connector runs the execution, retrieves the BUGS output (in the form of CODA files), which is then automatically converted into the Standard Output (SO) format by a BUGS to SO output converter. Finally, the connector retrieves the SO file, which becomes available for the user to perform graphical convergence diagnostics and posterior inference.

## A.2. Methods

### A.2.1. Software

The main software modules developed in this work are represented in Figure A.1 (with red boxes), and a detailed description of each of them is reported in *Supplementary Methods to Appendix A*. The version of IOF including the WinBUGS plugin (v.2.0) used is freely available at <http://aimed11.unipv.it/DDMoReIOF+WinBUGSplugin2.0/>, while a previous version of the plugin (v.1.0) is integrated into the official IOF public release (134).

### A.2.2. Implemented example workflow overview

A complex workflow, involving two diabetes-related published models (146, 148), has been executed within the IOF and is here proposed as an advanced real-world case study.

In the diabetic research area, it is of crucial importance to assess the insulin response to a glucose stimulus to understand the  $\beta$ -cell function in pathological states (144,145). The intravenous glucose tolerance test (IVGTT) is one of the simplest experiments to do that. To assess insulin response from IVGTT data, the insulin minimal model (MM) is widely used

---

(144–146), but it requires the knowledge of the individual C-peptide (CP) kinetics, which is described in the literature by a linear two-compartment model [45]. This model assumes that CP is secreted into the central compartment (compartment 1), from which it is eliminated or it is distributed into the peripheral one (compartment 2). Therefore, CP kinetics is fully characterized by four parameters:  $k_{01}$ ,  $k_{21}$ ,  $k_{12}$ , and  $V$ , where  $k_{ij}$  is the transfer rate from compartment  $j$  to  $i$  and  $V$  is the central compartment volume. The four parameters in a given individual can be estimated from the knowledge of age, sex, BSA, and health condition (normal, obese, diabetic), using a linear regression model and nonlinear algebraic relationships (147,148) (see paragraph A.2.3).

All these steps have been implemented following three approaches, based on different ways to propagate uncertainty (Figure A.2). The main steps, illustrated in detail in Figure A.2, include:

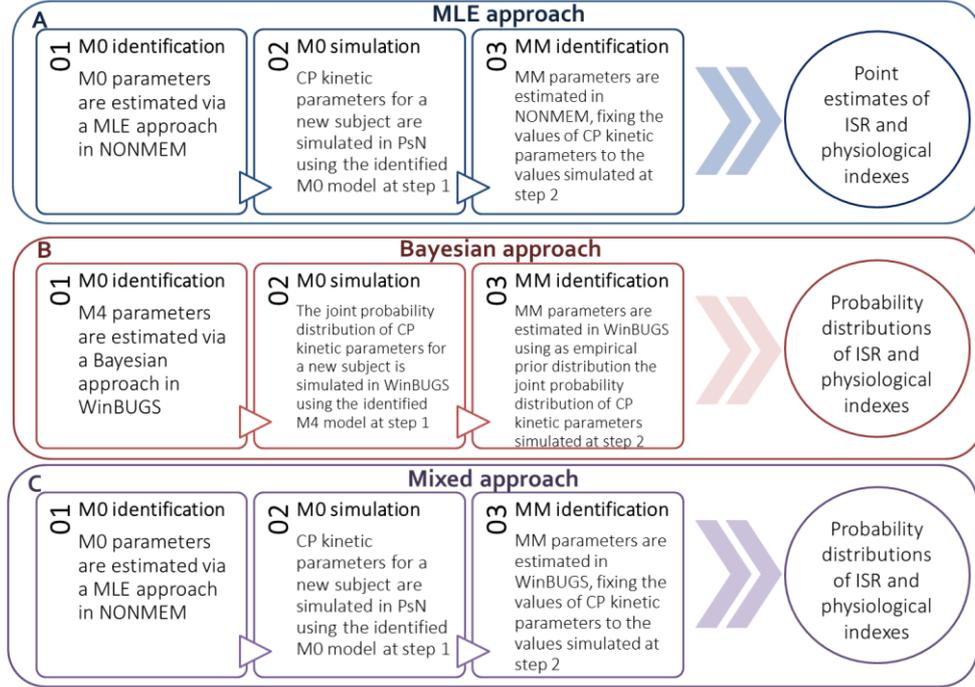
1. Identification of the population regression model from a large dataset of CP kinetic model parameters (see paragraph A.2.3).
2. Estimation of the CP kinetic parameters of a new subject, by using the identified population model.
3. Estimation of insulin secretion rate (ISR) and physiological indexes (e.g.  $\beta$ -cell sensitivity) by identifying the MM, using the CP kinetic parameters obtained above, and CP and glucose plasma concentration data of the new subject, coming from an IVGTT.

Approach 1 (maximum likelihood estimation - MLE - approach, Figure A.2-A) aims to obtain point-estimates of the variables of interest, without propagating parameter uncertainty throughout the steps 1-3. In this case, the point-estimates of the CP kinetic parameters, obtained at step 2, are used as fixed parameters of the MM at step 3.

Approach 2 (Bayesian approach, Figure A.2-B) aims to provide a statistical framework to properly handle the uncertainty and propagate it through all the workflow steps. In this approach, all the model elements (i.e., data, parameters, errors) are stochastic variables described by probability distributions. Therefore, the joint distribution of the CP kinetic parameters (obtained at step 2) is used as the prior distribution of these parameters in the MM.

Finally, Approach 3 (mixed approach, Figure A.2-C) includes the identification of the MM (step 3) via Bayesian approach but fixing the CP kinetic parameters to the values obtained from step 2 via Approach 1.

The software tools used via IOF to carry out the described tasks are NONMEM v.7.3, PsN v.4.4.8, Xpose v.4.5.3 and WinBUGS v.1.4 (with BlackBox v.1.5 (121) and the BUGSModelLibrary v.1.2).



**Figure A.2:** Modeling approaches implemented in this study. The following number of chains, burn-in iterations, updates and thin were used. M4 identification (Bayesian approach): 1 chain, 1,000 burn-in iteration, 100,000 updates, thin=10, repeated for three times, each time using the last values of the chain as initial values for all the population parameters, to eventually obtain 300,000 chain samples; insulin minimal model (MM) identification (Bayesian approach): 1 chain, 1,000 burn-in iterations, 20,000 updates, thin=1; MM identification (mixed approach): 1 chain, 1,000 burn-in iterations, 80,000 updates, thin=5.

### A.2.3 Mathematical models

#### *A population regression model to estimate CP kinetic parameters*

As reported in (148), the four parameters of the compartment model, i.e.  $k_{01}$ ,  $k_{21}$ ,  $k_{12}$  and  $V$ , can be obtained from the following macro constants: short half-life ( $t_s$ ), long half-life ( $t_l$ ), amplitude fraction ( $F$ ) and volume of distribution ( $V$ ), using the algebraic equations below:

$$k_{12} = \ln(2) \left( \frac{F}{t_l} + \frac{1-F}{t_s} \right), \quad (5.1)$$

$$k_{01} = \left( \frac{\ln(2)}{t_s} \right) \left( \frac{\ln(2)}{t_l} \right) \left( \frac{1}{k_{12}} \right), \quad (5.2)$$

$$k_{21} = \left( \frac{\ln(2)}{t_s} \right) + \left( \frac{\ln(2)}{t_l} \right) - k_{12} - k_{01}. \quad (5.3)$$

The four macro constants, in turn, can be derived in each subject via four linear regression models:

$$\theta_{ts_i} = \begin{cases} m_{tsn} \text{ if } HSTATUS_i \text{ is normal,} \\ m_{tso} \text{ if } HSTATUS_i \text{ is obese,} \\ m_{tsd} \text{ if } HSTATUS_i \text{ is diabetic,} \end{cases} \quad (5.4)$$

$$\theta_{F_i} = \begin{cases} m_{Fn} \text{ if } HSTATUS_i \text{ is normal,} \\ m_{Fo} \text{ if } HSTATUS_i \text{ is obese,} \\ m_{Fd} \text{ if } HSTATUS_i \text{ is diabetic,} \end{cases} \quad (5.5)$$

$$\theta_{V_i} = \begin{cases} a_{Vm} + b_{Vm} \cdot BSA_i \text{ if } SEX_i \text{ is male,} \\ a_{Vf} + b_{Vf} \cdot BSA_i \text{ if } SEX_i \text{ is female,} \end{cases} \quad (5.6)$$

$$\theta_{tl_i} = a_{tl} + b_{tl} \cdot AGE_i, \quad (5.7)$$

where AGE is in years and BSA, expressed in  $m^2$ , is calculated as  $0.20247 \times \text{Height}(m)^{0.725} \times \text{Weight}(kg)^{0.425}$ . For sake of simplicity, we will denote the four regression models as:

$$\boldsymbol{\theta}_i = \mathbf{f}(\tilde{\boldsymbol{\theta}}, \mathbf{U}_i), \quad (5.8)$$

where  $\boldsymbol{\theta}_i = [\theta_{ts_i}, \theta_{F_i}, \theta_{V_i}, \theta_{tl_i}]$  is the vector of the predictions of the four regressions for the  $i$ -th individual,  $\mathbf{U}_i = [HSTATUS_i, SEX_i, BSA_i, AGE_i]$  is the vector of individual covariates, and  $\tilde{\boldsymbol{\theta}} = [m_{tsn}, m_{tso}, m_{tsd}, m_{Fn}, m_{Fo}, m_{Fd}, a_{Vm}, b_{Vm}, a_{Vf}, b_{Vf}, a_{tl}, b_{tl}]$  is the vector of population parameters.

IIV of model parameters was assumed to be normally distributed. The individual macro constants  $\boldsymbol{\varphi}_i = [ts_i, F_i, V_i, tl_i]$  for the  $i$ -th individual are therefore calculated as:

$$\boldsymbol{\varphi}_i = \boldsymbol{\theta}_i + \boldsymbol{\eta}_i, \quad (5.9)$$

where  $\boldsymbol{\eta}_i = [\eta_{ts_i}, \eta_{F_i}, \eta_{V_i}, \eta_{tl_i}]$  is the vector of random effects which accounts for the IIV. Assuming independent random effects with unknown variance, the described model is equivalent to the one proposed by Van Cauter et al. (147). This model version will be referred to as M0.

Following (148), when correlations exist between the elements of  $\boldsymbol{\eta}_i$ , while they are independent between different subjects, we have:

$$\boldsymbol{\eta}_i \sim N(0, \boldsymbol{\Sigma}), \quad (5.10)$$

where  $\boldsymbol{\Sigma}$  is the full (4x4) covariance matrix.

To define a Bayesian model (147), priors on  $\tilde{\boldsymbol{\theta}}$  and on  $\boldsymbol{\Sigma}^{-1}$  are specified:

$$\tilde{\boldsymbol{\theta}} \sim N(\tilde{\boldsymbol{\theta}}_0, \boldsymbol{\Sigma}_0^{-1}), \quad (5.11)$$

$$\Sigma^{-1} \sim W(\rho, \mathbf{R}), \quad (5.12)$$

where  $\tilde{\theta}_0, \Sigma_0, \rho, \mathbf{R}$  are fixed prior parameters and  $W$  is the Wishart distribution with mean  $\rho \cdot \mathbf{R}$ . The following values were chosen according to the original publication (148):  $\tilde{\theta}_0 = [5 \ 5 \ 5 \ 1 \ 1 \ 1 \ 30 \ 1 \ 1 \ 1 \ 1 \ 1]$ ,  $\Sigma_0$  is a diagonal matrix with  $\tilde{\theta}_0$  square elements on the diagonal,  $\rho = 10$ , and  $\mathbf{R} = \rho^{-1}(0.01 \cdot \text{diag}([5 \ 1 \ 30 \ 4]))^{-1}$ . This Bayesian model version will be referred to as M4.

The MDL code of the described models is freely available for downloading in the DDMoRe Model Repository (143) at <http://repository.ddmore.eu/model/DDMODEL00000110>.

### ***Glucose-insulin minimal model***

The MM consists of two systems of differential equations, describing CP kinetics and ISR after a glucose perturbation (e.g. IVGTT), respectively (144,146).

The first subsystem is composed by the following equations:

$$\begin{cases} \frac{dCP_1(t)}{dt} = -(k_{01} + k_{21})CP_1(t) + k_{12}CP_2(t) + ISR(t) \\ \frac{dCP_2(t)}{dt} = -k_{12}CP_2(t) + k_{21}CP_1(t) \end{cases} \quad (5.13)$$

$$CP_1(0) = CP_2(0) = 0$$

where  $CP_1(t)$  and  $CP_2(t)$  are the CP concentration ( $\text{pmol} \cdot \text{l}^{-1}$ ) in compartment 1 and 2, respectively, and  $ISR(t)$  ( $\text{pmol} \cdot \text{l}^{-1} \cdot \text{min}^{-1}$ ) is the insulin (and therefore CP) secretion rate expressed as deviation from the basal and normalized by the volume of compartment 1 (V).

The second subsystem is composed by the following equations:

$$ISR(t) = mX(t) \quad (5.14)$$

$$\frac{dX(t)}{dt} = -ISR(t) + Y(t) \quad (5.15)$$

$$\frac{dY(t)}{dt} = \begin{cases} -\alpha(Y(t) - \beta(G(t) - h)) & \text{if } G(t) > h \\ -\alpha Y(t) & \text{otherwise} \end{cases} \quad (5.16)$$

$$X(0) = x_0, Y(0) = 0$$

where  $X(t)$  ( $\text{pmol} \cdot \text{l}^{-1}$ ) represents the concentration of CP in  $\beta$ -cells,  $m$  ( $\text{min}^{-1}$ ) represents the proportionality constant relating CP concentration in  $\beta$ -cells to insulin secretion rate, and  $Y(t)$  ( $\text{pmol} \cdot \text{l}^{-1} \cdot \text{min}^{-1}$ ) is a provisional factor stimulated when glucose plasma concentration is above the threshold  $h$  ( $\text{pmol} \cdot \text{l}^{-1}$ ). The initial condition  $X(0)=x_0$  ( $\text{pmol} \cdot \text{l}^{-1}$ ) represents the amount

of insulin secreted as an impulse in response to the elevated glucose level after the bolus. This first-phase is followed by a slower second phase governed by the provisional factor  $Y(t)$ , which tends to reach, with a time constant  $1/\alpha$  (min), a steady-state value linearly related, via parameter  $\beta$  ( $\text{min}^{-1}$ ), to the glucose concentration  $G(t)$  above the threshold value  $h$ .

The MM parameters of the first subsystem are:  $k_{01}$  ( $\text{min}^{-1}$ ),  $k_{12}$  ( $\text{min}^{-1}$ ), and  $k_{21}$  ( $\text{min}^{-1}$ ), illustrated above; they can be fixed to the values obtained via M0 or estimated via Bayesian approach. The MM parameters of the second subsystem are:  $h$  ( $\text{pmol}\cdot\text{l}^{-1}$ ),  $x_0$  ( $\text{pmol}\cdot\text{l}^{-1}$ ),  $\beta$  ( $\text{min}^{-1}$ ),  $m$  ( $\text{min}^{-1}$ ), and  $\alpha$  ( $\text{min}^{-1}$ ). CP and glucose plasma concentrations are provided in the dataset; in particular, the model has CP plasma concentration as dependent variable and glucose concentration as time-varying covariate.

The residual error model was supposed normally distributed with mean 0 and constant CV fixed to 6%.

Two physiological indexes,  $\varphi_1$  and  $\varphi_2$ , characterizing  $\beta$ -cell sensitivity to glucose, are defined as:

$$\varphi_1 = \frac{x_0}{\Delta G}, \quad (5.17)$$

$$\varphi_2 = \beta, \quad (5.18)$$

where  $\Delta G$  ( $\text{pmol}\cdot\text{l}^{-1}$ ) is the maximum measured increment of the glucose plasma concentration after an IVGTT.

When estimating the MM via Bayesian approach, MM parameters are a priori assumed to be independent and normally distributed. An informative prior was chosen for the threshold  $h$  with mean equal to the basal glucose level and a CV of 3%. Weakly to moderate informative priors were assumed for the other MM parameters,  $x_0$ ,  $\beta$ ,  $m$ , and  $\alpha$ , with mean 1.8, 11, 0.06, and 0.5, respectively, and a CV of 100%.

The MDL model code is freely available for downloading at <http://repository.ddmore.eu/model/DDMODEL00000111>.

### ***Datasets***

The population regression model to estimate CP kinetic parameters was identified using a large dataset including information about health status, sex, age, BSA, and corresponding CP kinetics macro constants of 207 subjects (147,148).

The glucose-insulin minimal model was identified on glucose and CP plasma concentration data, obtained after an IVGTT experiment on a subject not included in the previous dataset (148,149) with the following covariates: normal health status, male, 25 years old, height 1.818 m and weight 70.7 kg.

## **A.3. Results**

---

### A.3.1. Workflow results using a MLE approach

Approach 1 was executed via IOF, using NONMEM and PsN as target tools for estimation and simulation, respectively (Figure A.2-A). Point-estimates and corresponding precisions of the M0 parameters are reported in Table A.1. All of them are identical to the values reported in the original publication, in which a different target tool (MATLAB) was used (148). Precisions of parameters obtained after bootstrapping via PsN (Figure A.S1) were consistent with the precisions reported in Table A.1. Continuous or categorical VPCs were also performed via PsN and Xpose (Figure A.S2).

The compartmental parameters of a new subject were calculated via PsN by simulating M0 with all its parameters fixed to their point-estimates and using the anthropometric parameters of the new subject. The resulting values of  $k_{01}$ ,  $k_{21}$  and  $k_{12}$  are reported in Table A.1 and are also identical to the ones obtained in the original work (148).

The MM was identified via NONMEM from IVGTT data of the new subject, with compartmental parameters fixed to the values obtained above. Point-estimates with their precisions and sensitivity indexes are reported in Table A.2. Xpose was used to generate diagnostic plots (not shown here). Finally, PsN was used to simulate the identified MM to obtain the predicted CP plasma concentration (Figure A.3-A) and ISR (Figure A.4) time course plots. Sensitivity indexes and ISR are consistent with the values obtained in the original work for this subject (146), although a direct comparison cannot be performed since in the mentioned work the MM was always tested in a Bayesian context.

Although this approach includes stochastic elements, such as IIV or residual error, all the model parameters are considered as deterministic elements during estimation and simulation, making it unsuitable for uncertainty propagation among different models and hierarchical levels. This task will be faced in the next sections, in which Bayesian approaches are adopted.

### A.3.2 Workflow results using a Bayesian approach

Approach 2 was executed using WinBUGS as target tool for estimation and simulation (Figure A.2-B). The posterior distribution of the M4 model population parameters was computed, and the relative point-estimates and uncertainties were derived (Table A.1). All of them are consistent with the values reported using a different target tool (MCMC implemented via MATLAB) (148). Trace plots (Figure A.S3), obtained via the *coda* R package, were used to assess Markov chain convergence to eventually set the burn-in. In this case, chains were highly correlated. For this reason, to reduce autocorrelation and to save disk space, a thin of 10 was chosen to give an effective number of independent samples of at least 500 for each parameter. To check the number of independent samples, the R function *effectiveSize* (available in the *coda* R package) was used. As it was carried

---

out for Approach 1, VPCs were also performed (Figure A.S4); in this case, a custom R function was used, relying on the computed simulation profiles.

Stochastic simulations of M4 were performed to obtain the compartmental parameters, and their precisions, for the new subject. In these simulations, the priors on model parameters were replaced with the joint posterior distribution obtained after model identification, as (sampled) empirical distribution on all the model parameters. The point estimates of the compartmental parameters (with their 95% confidence intervals) for the new subject, reported in Figure A.S5 and Table A.1, were highly consistent with the values of the original work (148).

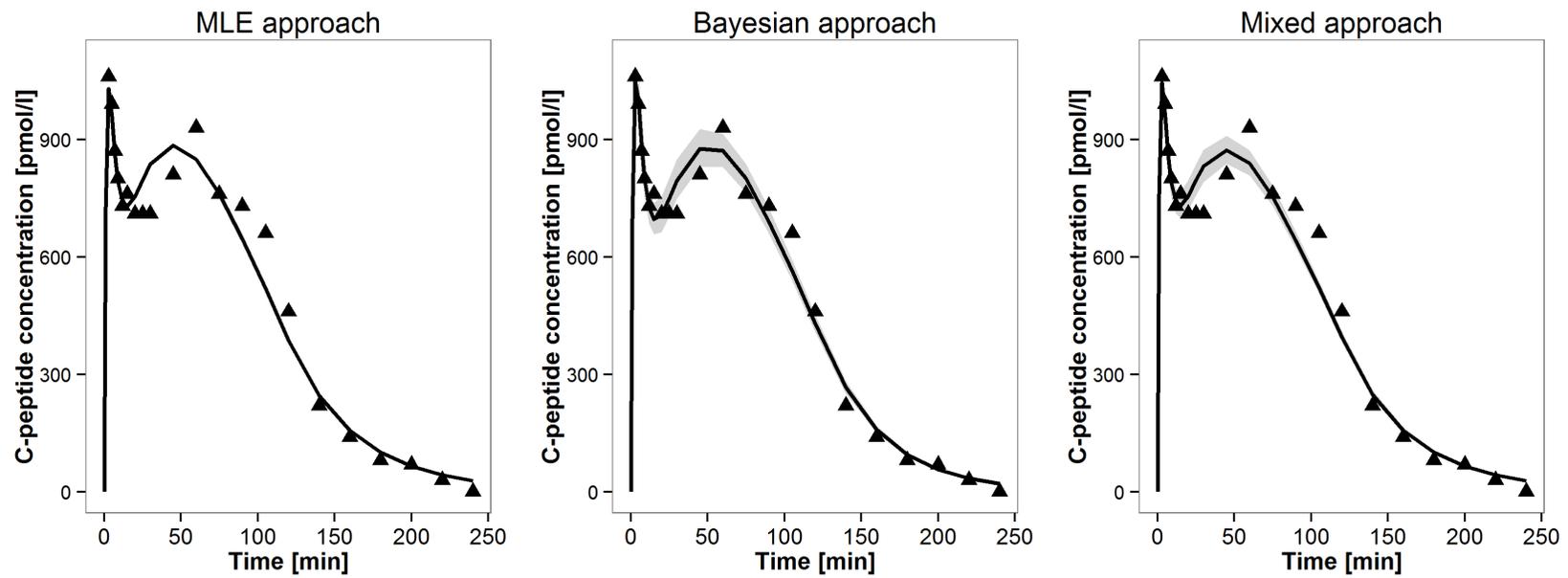
The joint probability distribution of the compartmental parameters of the new subject (500 samples) was used as empirical prior for these parameters during MM estimation. Posterior distributions were obtained for all the MM parameters and the relative point-estimates and uncertainties were computed (Table A.2), with results consistent with Approach 1 and the original publication (146).

The burn-in and thinning values were chosen via trace plot (not shown here) and *effectiveSize* function, as described above. The individual predicted vs. observed CP concentration plot is reported in Figure A.S6.

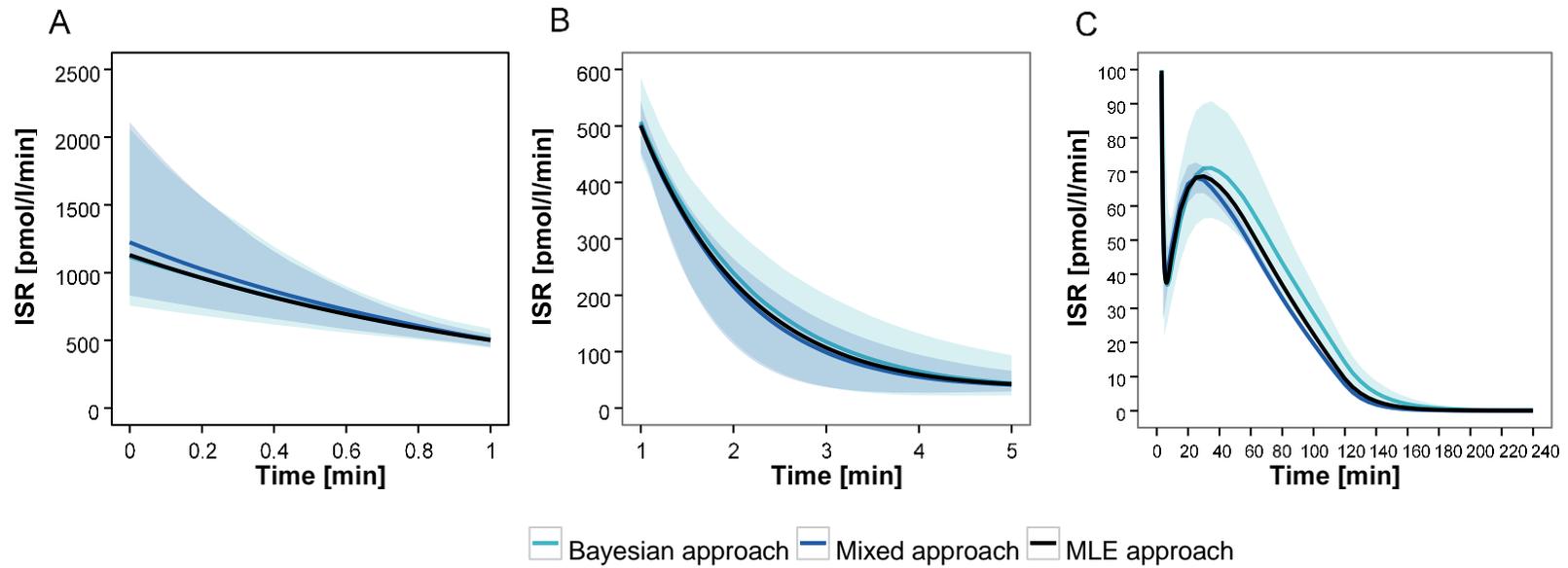
The identified MM was simulated to obtain the predicted CP plasma concentration (Figure A.3-B) and ISR (Figure A.4) time course plots. As before, sensitivity indexes and ISR are consistent with the values obtained in the original work for this subject (146).

### **A.3.3. Workflow results using a mixed approach**

Approach 3 is a combination of Approach 1 and 2 (Figure A.2-C). The compartmental parameters simulated with NONMEM in Approach 1 (Table A.1) were used as fixed values during MM Bayesian estimation with WinBUGS. Point-estimates and uncertainty of the MM parameters and sensitivity indexes in the new subject are reported in Table A.1. As before, all of them were consistent with the values estimated in Approach 1 and 2, as well as the ones in the original publication (146). As in Approach 2, predicted CP and ISR were obtained via stochastic simulations, reported in Figure A.3-C and Figure A.4. The numbers of burn-in iterations, update iterations and thin were chosen as described above (trace plots not shown here).



**Figure A.3:** Plot of individual fits for the insulin minimal model (MM). The panels correspond to the three proposed approaches: maximum likelihood estimation (MLE) approach (panel A), Bayesian approach (panel B), and mixed approach (panel C). The triangles represent individual C-peptide concentration data, the solid line represents the model fit line, the shaded grey area depicts the 95% confidence interval calculated without considering residual error variability.



**Figure A.4:** The reconstructed insulin secretion rate (ISR) profile. Expected time course (solid line) and 95% confidence interval (shaded area) of the ISR estimated via the three implemented approaches. Panels A, B, and C, correspond to time intervals [0,1], [1,5], [5,240] (min), respectively.

**Table A.1:** Final parameter estimates for M0 and M4.  
RSE: relative standard error; CP: C-peptide.

<sup>a</sup>95% confidence interval of the posterior distribution,

<sup>b</sup>Elements of the full matrix  $\Omega$ .

<sup>c</sup>Standard deviations of the additive residual errors.

		M0		M4	
Parameter	Unit	Estimate	%RSE	Estimate	%RSE
Fixed effect					
m <sub>tsn</sub>	min	5.000	2.088	4.991	1.942
m <sub>tso</sub>	min	4.554	3.758	4.496	2.921
m <sub>tsd</sub>	min	4.594	3.727	4.693	2.957
m <sub>Fn</sub>	-	0.764	0.546	0.766	0.562
m <sub>Fo</sub>	-	0.782	0.629	0.781	0.784
m <sub>Fd</sub>	-	0.780	0.771	0.778	0.858
a <sub>tl</sub>	min	27.797	4.802	26.705	3.854
b <sub>tl</sub>	min/years	0.177	22.728	0.209	13.197
a <sub>Vm</sub>	L	0.495	181.584	0.344	131.067
b <sub>Vm</sub>	L/m <sup>2</sup>	1.982	22.630	2.061	10.730
a <sub>Vf</sub>	L	1.520	48.365	0.795	59.352
b <sub>Vf</sub>	L/m <sup>2</sup>	1.432	28.790	1.819	13.898
Random effect					
$\omega_{ts}^b$	min <sup>2</sup>	-	-	1.295	9.781
$\omega_F^b$	min <sup>2</sup>	-	-	0.002	9.822
$\omega_{tl}^b$	-	-	-	33.044	9.847
$\omega_V^b$	L <sup>2</sup>	-	-	0.713	9.796
$\omega_{ts,F}^b$	min	-	-	0.006	60.110
$\omega_{ts,tl}^b$	min <sup>2</sup>	-	-	3.250	15.488
$\omega_{ts,V}^b$	min·L	-	-	0.596	13.214
$\omega_{F,tl}^b$	min	-	-	0.071	26.966
$\omega_{F,V}^b$	L	-	-	-0.006	49.064
$\omega_{tl,V}^b$	min·L	-	-	1.915	19.095
$\sigma_{ADDts}^c$	-	1.143	5.394	-	-
$\sigma_{ADDf}^c$	-	0.041	5.625	-	-
$\sigma_{ADDtl}^c$	-	5.778	6.167	-	-
$\sigma_{ADDv}^c$	-	0.846	5.800	-	-
CP-kinetic parameters					
k <sub>01</sub>	min <sup>-1</sup>	0.061		0.062 (0.045-0.096) <sup>a</sup>	
k <sub>12</sub>	min <sup>-1</sup>	0.049		0.049 (0.032-0.085) <sup>a</sup>	
k <sub>21</sub>	min <sup>-1</sup>	0.050		0.049 (0.027-0.110) <sup>a</sup>	

**Table A.2:** Final parameter estimates for the insulin minimal model (MM) in the three approaches. %RSE for the MLE approach are not reported since covariance step using FOCE was not successful; the use of SAEM instead of FOCE was attempted but it did not provide reliable estimates due to convergence issues (data not shown).

RSE: relative standard error; MLE: maximum likelihood estimation; FOCE: first order conditional estimation; SAEM: stochastic approximation of expectation maximization.

		MLE approach		Bayesian approach		Mixed approach	
Parameter	Unit	Estimate	%RSE <sup>a</sup>	Estimate	%RSE	Estimate	%RSE
m	min <sup>-1</sup>	0.817	-	0.820	31.343	0.907	25.731
$\alpha$	min <sup>-1</sup>	0.061	-	0.051	11.863	0.074	7.149
$\beta$	min <sup>-1</sup>	9.790	-	10.498	12.254	8.909	3.139
x0	pmol/L	1.384	-	1.472	8.212	1.406	3.220
h	pmol/L	89.002	-	89.097	2.313	89.859	2.112
$\varphi_1$	-	81.966	-	87.177	-	83.314	-
$\varphi_2$	min <sup>-1</sup>	9.790	-	10.498	-	8.909	-

**Table A.3:** List of modeling features supported by the interoperability framework (IOF) with WinBUGS plugin.

Models including the reported features are specified as examples, referring to Modelling Description Language (MDL) files downloadable with the IOF public release (134) or from the model repository (143). Among all the modeling features supported by MDL, the limitations, if present, are also listed for each feature, explicitly specifying if they are due to the design of our software or to WinBUGS constraints (marked with the asterisk). Novel features which were not present in the public version final release of the IOF (134) are reported in bold.

ODE: ordinary differential equations; PK: pharmacokinetic.

Feature	Example	Limitations or assumptions
Multiple variability levels	UseCase1	Variability is supported on the population (prior), individual (between-subject) and observation (residual) levels
Algebraic and ODEs models	UseCase1, UseCase2	ODE models are solved via BUGSModelLibrary and WBDev. Initial time is always assumed to be zero.
IF-THEN-ELSE statements	UseCase1	
Univariate and multivariate distributions from ProbOnto knowledge-base	Table A.S2	Only the ones supported by WinBUGS*
Unary and binary operators	Table A.S1	Only the ones supported by WinBUGS*
Pairwise covariance and correlation encoding	UseCase1	

Feature	Example	Limitations or assumptions
Additive, proportional and combined observation error models	UseCase1	Only continuous outputs and structured expressions are supported: $OBSERVATION = PREDICTION + f*EPS$ , with $EPS$ distributed as a Normal and $f$ depending on the error model*
<b>Single, multiple independent observations and observations with multivariate distribution</b>	<b>M4</b>	<b>Only univariate and multivariate Normal distributions are supported</b>
Multiple dosing	Simeoni2004	
Multiple administration routes	UseCase4_1	
NONMEM-formatted data file	All models	Supported columns: ID, DV, DVID, EVID, MDV, AMT, CMT, SS, RATE, II, ADDL
Continuous and categorical covariates	UseCase5	All the covariates are assumed to be time-dependent; continuous ones are linearly interpolated; constant interpolation is performed on categorical ones
Transformation of covariates, individual parameters and observation models	UseCase5	
PK macros	UseCase5	
<b>Univariate/multivariate empirical and non-parametric prior distributions</b>	<b>MM</b>	<b>Sample(s) (for empirical) and bin(s)-probability (for non-parametric) values must be provided via external .csv file</b>
Structured (linear and general) expressions for individual parameters	UseCase1	
<b>Matrices and vectors</b>	<b>M4</b>	<b>They can only be used in population parameters and they cannot be used in IF-THEN-ELSE statements or ODEs*</b>
Structured (linear and general) expressions for individual parameters	UseCase1	

#### A.3.4. Supported features of the WinBUGS plugin

Considering the models and performed tasks in Approach 2, the support of a wide number of features was demonstrated, including: estimation and simulation tasks, population and single-subject models, algebraic and ODE structural models, continuous and categorical covariates, time-dependent forcing functions, multiple observations, IF-THEN-ELSE statements, correlated random effects, parametric and non-parametric (i.e. expressed with frequency table) or empirical (i.e. expressed through a list of samples) prior distributions. Specifically, our software plugin supports all the WinBUGS-compatible probability distributions included in the ProbOnto ontology (v.2.0) (150), which is used as a standard knowledge-base in MDL and PharmML.

In addition to M4 and MM, the WinBUGS plugin was tested on a collection of about 200 additional models, including different features of interest in pharmacometrics. Based on these tests, a full list of supported features is reported in Table A.3. Compared to the WinBUGS plugin

---

available in the IOF public release (134), in the updated plugin the support of new features (highlighted in Table A.3) has been added.

Users can retrieve the IOF with the WinBUGS plugin (v.2.0) used by downloading it from <http://aimed11.unipv.it/DDMoReIOF+WinBUGSplugin2.0/>.

## A.4. Discussion

Reproducibility and interoperability of models code among different target languages and software tools have been demonstrated with a complex workflow, which has been executed via three different approaches combining several tools.

With Approach 1 we have demonstrated how, in a single R script, users can estimate model parameters using NONMEM and do model qualification using PsN and Xpose. Although all the steps of the workflow could have been performed with the stand-alone versions of these software tools, writing an unbroken R script can significantly support the error-free reproducibility of the carried out analysis, since all the task-implementing commands can be included in a single file (e.g., fixing model parameters to the previously estimated values, reuse the same model for estimation, simulation and VPC). Moreover, the output of all the performed tasks was saved in a SO file, supporting subsequent result comparisons among different target tools, e.g. through the application of standard Xpose graphic functions.

The other key advantage of the IOF features is interoperability, which allows reusing the same data and/or model encoding, for executing the desired tasks via different target languages and tools. This feature has been demonstrated in Approach 2 and 3, in which the code (model and data) used in Approach 1 has been easily reused in a Bayesian setting, thanks to the modular structure of the MDL. In fact, MDL objects can be grouped in different ways within IOF to execute different tasks, even in different software tools.

Further, with Approach 2 and 3, we have demonstrated the possibility of executing Bayesian estimation and stochastic simulation tasks in the IOF via WinBUGS by using only MDL and R scripts. Both the estimation and simulation results were stored in a SO file, like in Approach 1. Therefore, interoperability and standardization have also been highlighted at the level of result management by the possibility of applying, in all the tested approaches, “universal” functions to SO files, without taking into account which software generated results and which was the specific format for storing them. That was made possible thanks to the standardized format of the IOF outputs (i.e. SO) that promotes interoperability and enables direct comparisons among results coming from different target tools and tasks. For instance, an easy to do comparison between the confidence bands around the ISR estimates (Figure A.4) highlights, in this case, that the central tendency of the reconstructed ISR time course is comparable among the different approaches (with only a slight overestimation of the ISR in the MLE

---

approach), while the propagated variability, represented by the 95% confidence interval around the central tendency, increases from the mixed approach to the Bayesian approach, as expected. It is worth noting that the confidence region of the mixed approach is completely included within the one of the Bayesian approach only after the first 5 minutes, while in the first two phases the differences between the confidence regions are quite limited.

Finally, the integration of WinBUGS within the DDMoRe IOF, through the WinBUGS plugin, has even reduced the complexity of directly implementing PKPD models, which are normally encoded in the stand-alone version of WinBUGS by combining in a not trivial way BUGS code and Component Pascal languages. In particular, modelers can encode the desired MDL files by writing them from scratch, taking advantage of the MDL-IDE and a detailed user guide, or modifying existing model files, such as the ones available in the DDMoRe Model Repository. The modular structure of MDL facilitates such task, by enabling the re-use of blocks from existing model files, which may be modified only in terms of, e.g., data or prior distribution of specific parameters. Any MDL file can be executed via user-defined R scripts, like the ones programmed in this work. Once executed via specific R functions, MDL can successfully serve as a BUGS/Component Pascal translation system, since every execution creates and makes available to users all the model files in the target code. If needed, the resulting files can be integrated, modified and executed by users in stand-alone WinBUGS.

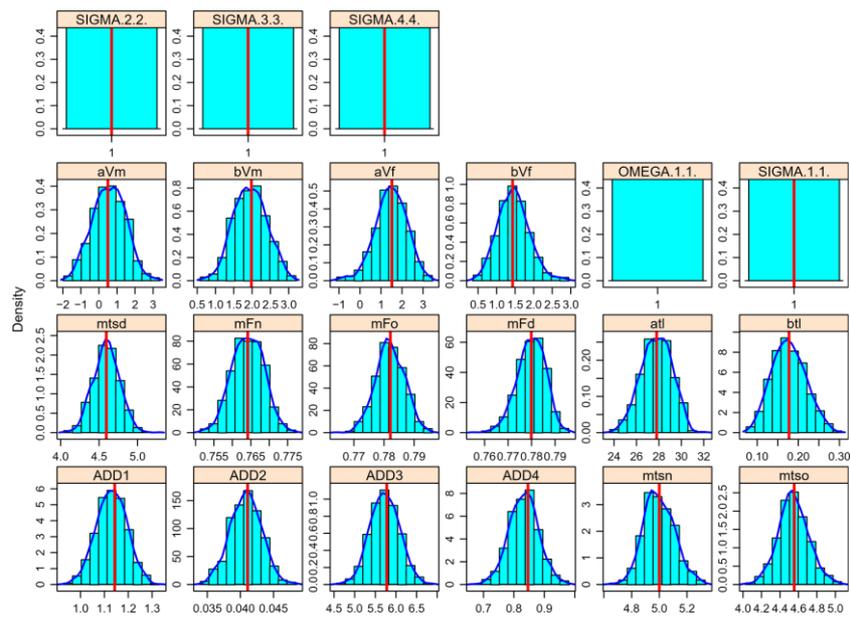
The work herein presented relied on the WinBUGS plugin available functionality and extensibility; however, the lessons learned could facilitate the development of future DDMoRe IOF plugins for other promising Bayesian tools, such as Stan.

The plugin used in this work is publicly available and supports a plethora of pharmacometric modelling features (Table A.3). It is expected to significantly facilitate Bayesian model encoding, execution, and results comparison among different estimation/simulation tools, by addressing reproducibility and interoperability, two long-standing problems in pharmacometric modelling, as well as by “making easy” the encoding of PKPD models in WinBUGS. This can substantially contribute to boosting the adoption of the Bayesian approach in pharmacometrics.

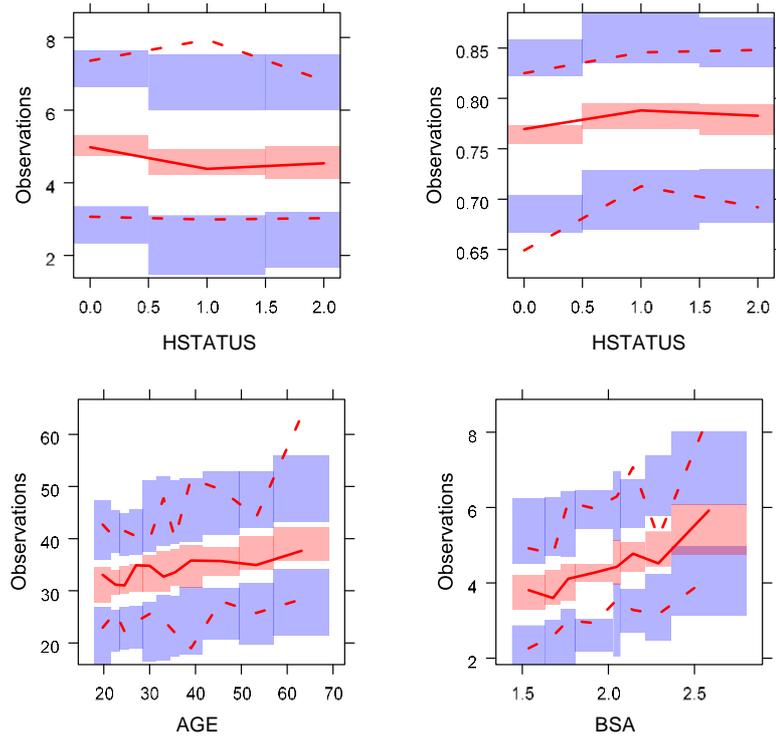
---

# Supplementary material to Appendix A

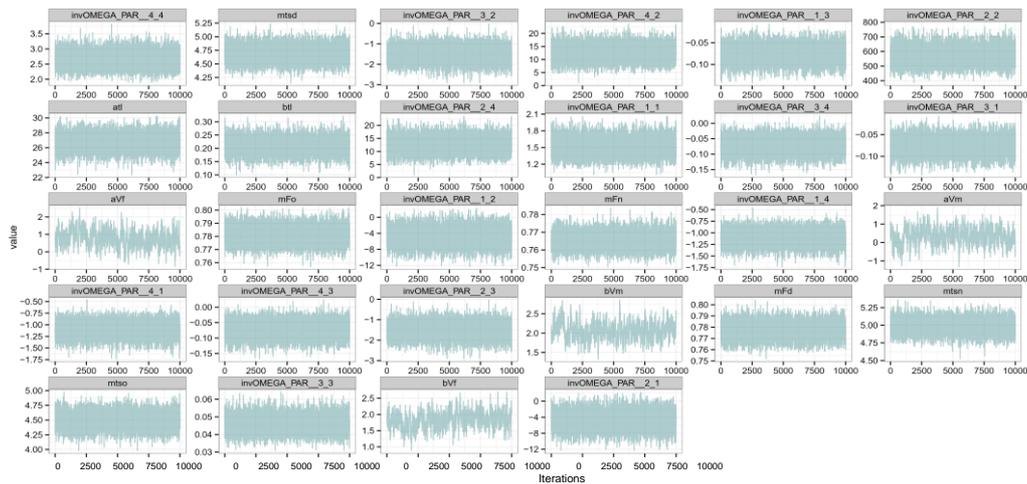
---



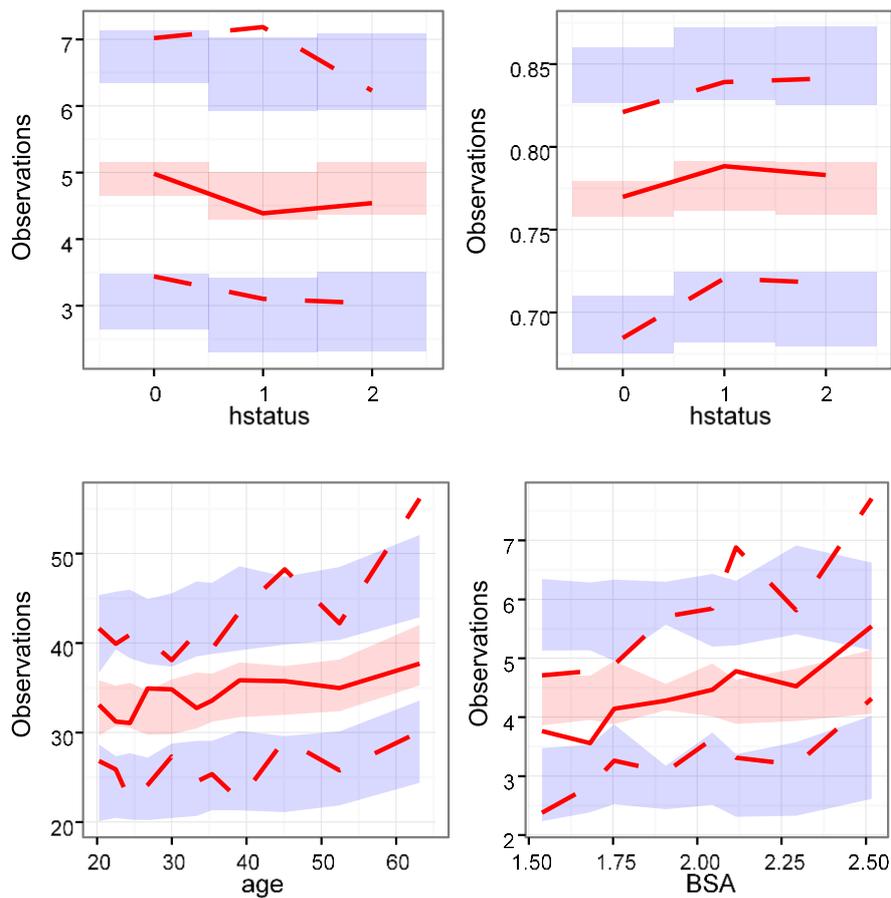
**Figure A.S1:** Histograms of bootstrap M0 parameter estimates with normal density curves obtained via the maximum likelihood estimation (MLE) approach.



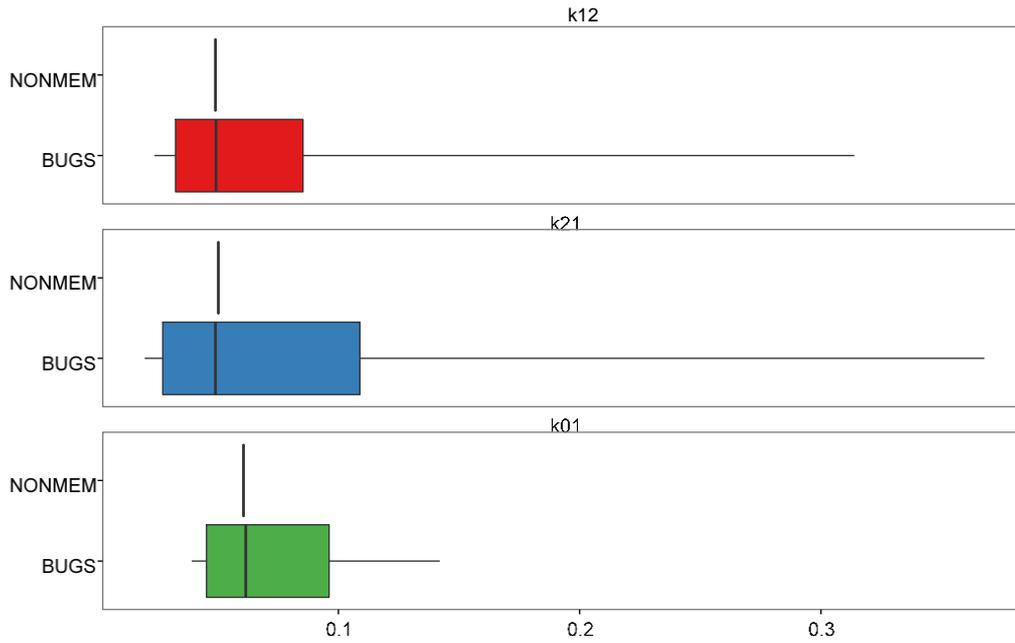
**Figure A.S2:** Visual predictive check (VPC) plots of M0 obtained via the maximum likelihood estimation (MLE) approach. Red solid line, median of observed data; red dashed line, 95% confidence interval (CI) of observed data; red shaded area, 95% CI around median of simulated data; blue shaded area, 95% CI around 95% prediction interval (PI) of simulated data for short half-life (top-left panel), amplitude fraction (top-right panel), long half-life (bottom-left panel), and volume of distribution (bottom-right panel).



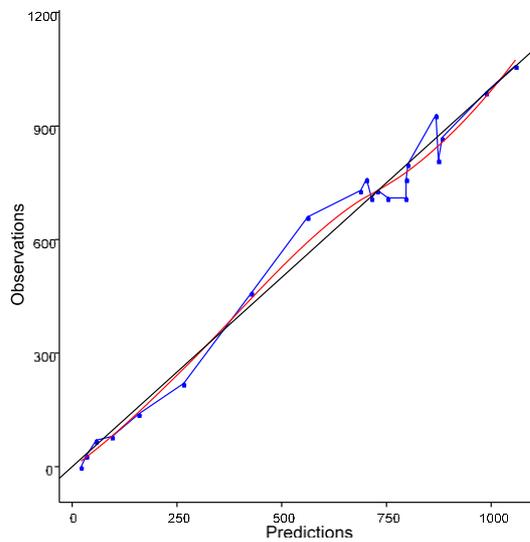
**Figure A.S3:** Trace plots of M4 obtained via the Bayesian approach.



**Figure A.S4:** Visual predictive check (VPC) plots of M4 obtained via the Bayesian approach. Red solid line, median of observed data; red dashed line, 95% confidence interval (CI) of observed data; red shaded area, 95% CI around median of simulated data; blue shaded area, 95% CI around 95% prediction interval (PI) of simulated data for short half-life (top-left panel), amplitude fraction (top-right panel), long half-life (bottom-left), and volume of distribution (bottom-right panel).



**Figure A.S5:** C-peptide (CP) kinetic parameters for a new subject. The point estimates obtained via the maximum likelihood estimation (MLE) approach are denoted by a solid line and the 95% confidence intervals obtained via the Bayesian approach by a colored zone (with the median denoted by a solid line).



**Figure A.S6:** Observations vs. individual predictions of the final insulin minimal model (MM) obtained via the Bayesian approach.

---

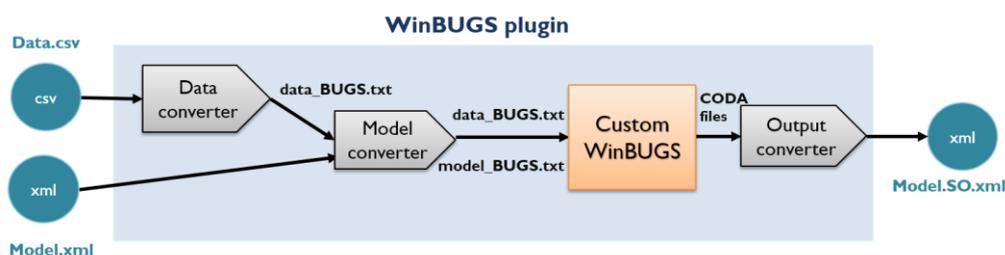
## Supplementary methods to Appendix A

---

### Software implementation: WinBUGS plugin and connector

The WinBUGS plugin includes the whole environment (converters, tools, and scripts) necessary to automatically translate and execute a PharmML model with WinBUGS as target tool in the DDMoRe IOF. The plugin receives as inputs the PharmML model and the related data file and provides as final output the results in SO format. The WinBUGS connector controls the entire model execution workflow.

The input and output files are represented in Figure A.S7 as blue circles, while the light blue box includes the main components of the plugin. The overall model estimation/simulation process is accomplished by the following modules: the data file converter (NONMEM-to-BUGS format translator), the model converter (PharmML-to-WinBUGS format translator), the WinBUGS software (a custom version of WinBUGS including all the required extensions purposely built for executing the specific model) and the output converter (CODA-to-SO format translator).



**Figure A.S7:** A high-level processing schema of a model execution flow for estimation or simulation tasks with the WinBUGS plugin.

#### *NONMEM-formatted to BUGS data file converter*

This module takes as input the NONMEM-formatted data file and translates it into BUGS format. The converter is written in R programming language.

---

### ***WinBUGS converter***

The main function of the WinBUGS converter is the translation of the model written in PharmML v.0.8.1 into BUGS code. However, as the BUGS language does not support some relevant features, like conditional statements, and WinBUGS is distributed with an ODE solver affected by relevant limitations, the converter generates a set of additional Pascal and script files necessary to support the execution of a wider range of complex models (Table A.3 for the complete list of supported features).

The list and the usage of these files are explained below.

Another important task of the converter is to update the data file to remove unused variables and remap multiple observations in a suitable matrix format, given the information encoded in the PharmML model. The implementation language of the converter is Java 1.7. The converter is based on several libraries:

- libPharmML v.0.7.3-1 (library providing PharmML to Java Objects mapping), libPharmML-PKMacro v.0.3.2-1 (library adding PharmML PK-macro support to libPharmML), provided by the DDMoRe consortium.
- CCoPI-Mono v.0.0.8, library providing generic PharmML parsing and converter functionality, developed by Cyprotex Discovery Ltd. The library provides binary-tree representations of equations, unary and binary operations and piecewise (conditional) statements. The library also defines Service Provider Interface (SPI) to be implemented by the language converters, to reuse (or customize) common PharmML processing and data extraction logic. The library is open source available at <https://sourceforge.net/projects/ccopi-mono/>.

### ***Custom WinBUGS version***

The version of WinBUGS, which is built and executed at each model run by the plugin, is based on the WinBUGS v.1.4 (107) distribution extended via WBDDev (117) with custom functions based on the BUGSModelLibrary v.1.2 (118,119). The use of a customized version of WinBUGS integrating new functions is necessary to support many relevant features reported in Table A.3, for instance conditional statements. The steps and the mechanisms underlining the generation of this release of WinBUGS is controlled by the WinBUGS connector, as described in the previous sections, and exploit the BlackBox v.1.5 environment capabilities (121).

### ***Output converter***

The output converter translates the CODA files, containing the results of model execution, into SO format (v.0.3.1). The CODA format consists of two ASCII files with extensions *.ind* and *.out*, respectively. The *.ind* file contains the parameter names and indexes necessary to correctly read the *.out* file, which contains the samples of the posterior distributions generated by the MCMC algorithm. The output converter generates an XML file

---

containing summaries and statistics of the results and a set of external *.csv* files, containing the samples. The implementation language of the SO converter is Java v.1.7. The converter is based on the libPharmML-SO v.0.4.2-1 Java library.

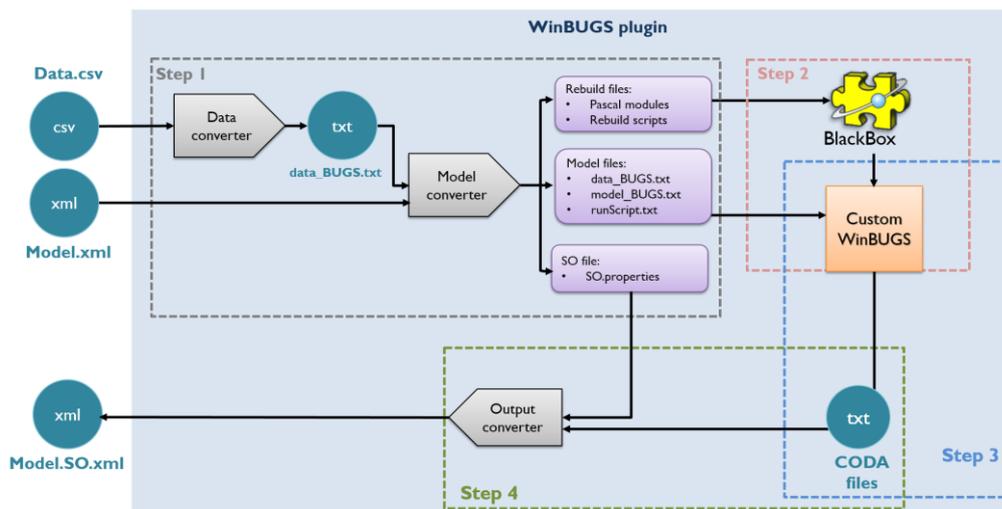
### ***WinBUGS connector and rebuild process***

As anticipated above, the WinBUGS version used to execute a task on a Bayesian model is a customized version of the WinBUGS v.1.4 distribution, automatically generated by the WinBUGS connector, in order to support the required features reported in Table A.3. The automatic rebuild of WinBUGS exploits WDev, the Development Interface designed to enable WinBUGS v.1.4 users to implement their own specialized functions 'hard-wiring' them into the system via Pascal components. For this purpose, in addition to the BUGS model, the converter generates a set of model-specific Pascal components (referred hereafter as “Pascal model library”) to be compiled and linked to WinBUGS core distribution and the scripts, necessary to perform the rebuild. All the tools necessary to this step are included in the plugin.

The diagram in Figure A.S8 shows the detailed sequence of actions carried out by the WinBUGS connector to convert a model, rebuild WinBUGS, and execute a specific task on the converted model.

They are split into four sequential steps:

- Step 1 (Model/script files generation): the WinBUGS converter is run to generate the files and scripts necessary for executing steps 2-4;
- Step 2 (WinBUGS rebuilding): the connector rebuilds WinBUGS by hard-wiring the Pascal model library;
- Step 3 (Model execution): execution of the model with the custom version of WinBUGS generated at Step 2;
- Step 4 (SO generation): results conversion to SO format.



**Figure A.S8:** Schema of the sequence of actions performed by the WinBUGS connector to convert a model, rebuild WinBUGS, and execute a specific task on the converted model.

### Step 1

Starting from the input PharmML model file and the converted BUGS data file, the WinBUGS converter generates the following files which are then used by the connector in the next steps (Figure A.S8):

- Rebuild files:
  - Model-specific Pascal components (i.e. “Pascal model library”) necessary to embed the new required functionalities, like piecewise functions, categorical covariate interpolation, and differential equation solvers, into WinBUGS;
  - script files necessary to rebuild WinBUGS. Such files will be provided as input to BlackBox Component Builder v.1.5, the build engine used to automate the process for building WinBUGS releases by hard-wiring new functions. This step is mandatory for each model run.

All these files are the input to Step 2.

- Model files:
  - BUGS-model (named *model\_BUGS.txt*);
  - Updated data file (named *data\_BUGS.txt*);
  - Script for batch execution of the BUGS model (*runScript.txt*).

These files are the input to Step 3.

- SO file:
  - properties file (named *SO.properties*), containing the information required by the WinBUGS SO converter.

This file is the input to Step 4.

### Step 2

This step consists of the building of a version of WinBUGS tailored to the specific model. The atomic operations carried out in this step by the

---

connector are: (a) preparation of the environment for rebuilding (e.g., symbolic links definition, folders creation, files moving into the appropriate folders); (b) conversion of the additional Pascal components (.txt files), generated by the WinBUGS converter, into the BlackBox internal format (.odc format); (c) custom WinBUGS generation by batch compilation and link of the modules using BlackBox.

### *Step 3*

In this step, the model is run in WinBUGS. The model features that cannot be directly represented in the BUGS language are implemented in specific Pascal functions hard-wired into WinBUGS and called by the BUGS model generated by the converter.

### *Step 4*

The output of WinBUGS is a set of CODA files that contain the results of the model estimation/simulation. The SO converter exploits the file named SO.properties to translate these data into the standard SO format. At the end of this step, the connector performs the environment cleaning and reset to the initial setup by rebuilding the standard version of WinBUGS ready for the next model execution.

## **Description and assumptions for the automatically generated BUGS model**

The WinBUGS data and model converters generate the following model files:

- *data\_BUGS.txt* (always)
- *modelBUGS.txt* (always)
- *ODEPascal.txt* + *PKModels.txt* + *ODEPascal\_Mod.txt* (only in case of ODE models)
- *FunctionPiecewise\*.txt* (only in case of conditional statements)
- *FunctionCovariate\*.txt* (only in case of categorical covariates)
- *RunScript.txt* (always)

### ***data\_BUGS.txt structure***

A NONMEM-formatted dataset follows specific requirements listed in Chapter 6 of the NONMEM Users Guide Part V - Introductory Guide, available at <https://nonmem.iconplc.com/nonmemVI/guides/V.pdf>. Starting from a NONMEM-formatted dataset in .csv format, the automatically translated *data\_BUGS.txt* file contains a named list (in the format required by BUGS) of the following elements:

- **N\_subj**: number of subjects;
- **N\_t**: a  $1 \times N\_subj$  vector, containing the number of grid points of each subject;

For each dependent variable:

- 
- **grid**: an  $N\_subj \times \max(N\_t)$  matrix, containing in the i-th row the grid points (specified in the TIME column) of the i-th subject (NA is used to fill the missing grid points among subjects);
  - **<ObservedVariableName>**: an  $N\_subj \times \max(N\_t)$  matrix, containing in the i-th row the values of the observed variable (specified in the DV column) of the i-th subject;

For each covariate (all the covariates are assumed to be time-dependent):

- **grid\_<CovariateName>**: an  $N\_subj \times \max\_m\_<CovariateName>$  matrix, containing in the i-th row the grid points of the TIME column, in which the covariate values relative to the i-th subject are specified;
- **<CovariateName>**: an  $N\_subj \times \max\_m\_<CovariateName>$  matrix, containing in the i-th row the values of the covariate relative to the i-th subject.

For a model defined through ODEs:

- **rate**: an  $N\_subj \times \max(N\_t)$  matrix, containing in the i-th row the values specified in the RATE column relative to the i-th subject. If the rate is specified without a dose, its value is set to 0 by default. If rate is missing, its value is set to 0 by default.
- **amt**: an  $N\_subj \times \max(N\_t)$  matrix, containing in the i-th row the values specified in the AMT (or DOSE) column relative to the i-th subject. If AMT is missing or NA, its value is set to 0.
- **ii**: an  $N\_subj \times \max(N\_t)$  matrix, containing in the i-th row the values specified in the II column relative to the i-th subject. If II is specified without a dose, its value will be ignored. If II column is missing, its value is set to 0 by default.
- **evid**: an  $N\_subj \times \max(N\_t)$  matrix, containing in the i-th row the values specified in the EVID column relative to the i-th subject. EVID values equal to 3 and 4 are ignored since they are not supported by WinBUGS.
- **cmt**: an  $N\_subj \times \max(N\_t)$  matrix, containing in the i-th row the values specified in the CMT column relative to the i-th subject. The numeric codes reported in the original dataset are mapped to the actual numbers of each derivative variable. If the AMT column is present but CMT column is missing, its value is set to 1 by default.
- **addl**: an  $N\_subj \times \max(N\_t)$  matrix, containing in the i-th row the values specified in the ADDL column relative to the i-th subject. If ADDL is specified without II and/or AMT or is missing, its value is set to 0 by default.
- **ss**: an  $N\_subj \times \max(N\_t)$  matrix, containing in the i-th row the values specified in the SS column relative to the i-th subject. If SS is specified without II or is missing, its value is set to 0 by default.

If covariates are used in the ODE model, the following elements are produced:

- **n\_cov\_cont**: number of continuous covariates;
- **n\_cov\_cat**: number of categorical covariates;

Then, for each covariate:

- 
- **max\_m\_<CovariateName>**: maximum number of grid points of the covariate among the different subjects;
  - **N\_t\_<CovariateName>**: vector of length equal to the number of grid points of the covariate of each subject;
  - **max\_m\_<CovariateName>**: maximum number of grid points of the covariate among the different subjects;

### *model\_BUGS.txt structure*

It contains at least two for cycles: the first loop in which *ind\_subj* is iterated over the different subjects, and the second loop in which *ind\_t* is iterated over the grid points of the *ind\_subj*-th subject. The same structure is also maintained when only one subject and/or only one grid point are/is present.

The following blocks of WinBUGS code are generated within the structure described above:

#### 1. *Covariates interpolation block*

Linear interpolation for continuous covariates (performed by *interp.function*) and piecewise constant interpolation for categorical covariates (performed by *interp.function.cost*) are here used to interpolate the covariates over the grid of the observed variable. The first grid point of the covariate cannot be greater than the first grid point of the observed variable. The last grid point of the covariate cannot be smaller than the last grid point of the observed variable.

#### 2. *Covariates transformation block*

Here the covariate transformation for both continuous and categorical covariates is specified. Conditional statements (see *Conditional statements management*) are here used to map the arbitrary numeric codes defined in the dataset for each categorical covariate with the appropriate values in 0-1 notation.

#### 3. *Individual variables definition block*

Here it is specified how the fixed effect variables (population parameters, covariates with their associated fixed effect parameters) and random effects combine to define the individual variables which will be used to calculate model predictions.

#### 4. *Model predictions block*

For an algebraic model, here model prediction is calculated using mathematical expressions that may involve population parameters and individual variables. For an ODE model, a call to a specific function is performed to calculate the numerical solution of ODEs (see ODE model management).

#### 5. *Likelihood definition block*

Here the probability distribution of the likelihood is defined in this form:  $Y_{ij} \sim N(Y_{mean_{ij}}, Y_{prec_{ij}})$ , where  $Y_{ij}$  is the *i*-th observation of the *j*-th subject,  $Y_{mean_{ij}}$  is the corresponding model prediction, and  $Y_{prec_{ij}}$  the corresponding precision (inverse of variance).

For example, in UseCase1 (downloadable with the IOF public release), where the warfarin population pharmacokinetic ODE model is

---

used (for a detailed description of the model see: <http://www.ddmore.eu/sites/ddmore/files/downloads/ReadMe.pdf> (page 4/30)), we have:

```
Y_mean[ind_subj,ind_t] <- (CC[ind_subj,ind_t] + 0)
Y_prec[ind_subj,ind_t] <- 1 / ((pow(((RUV_ADD + (RUV_PROP
*CC[ind_subj,ind_t])),2))*1)
Y[ind_subj,ind_t] ~ dnorm(Y_mean[ind_subj,ind_t], Y_prec[ind_subj,ind_t])
```

In case of multivariate probability distributions, the notation is slightly different and requires a specific MDL code (see *Multivariate observation model*).

#### 6. *Prior definition block*

Here the priors used in the Bayesian estimation are specified for all the parameters.

The main structure is summarized below.

```
model {
  for (ind_subj in 1:N_subj) {
    for (ind_t in 1:N_t[ind_subj]) {
      #time- and individual-dependent block
    }
    #individual-dependent block
  }
  #prior definition block
}
```

The time- and individual-dependent blocks always contain blocks I-II and IV-V, and, in case of individual parameters dependent on covariates, also block III is contained. The individual-dependent block contains block III in case of individual parameter not-dependent on covariates; the prior definition block includes block VI.

#### ***Nonparametric/Empirical prior distributions***

Sometimes it can be of interest to express prior distributions of model parameters without using canonical distributions, such as normal, lognormal, gamma, etc. In fact, a prior distribution can also be expressed by describing it through a frequency table (i.e. nonparametric distribution) or a set of samples (i.e. empirical distribution). In nonparametric distribution, the (joint) prior distribution is specified through a table (a .csv file) containing values-probability sets in the rows. If the probabilities are not specified, all the values are assumed to have equal probability (1/number of rows). In empirical distribution, the (joint) prior distribution is always specified through a table (a .csv file), in this case containing samples. In WinBUGS the implementation for nonparametric and empirical distributions is the same, for example:

```
ind_data_CL_V_multi ~ dcat(weight_CL_V[])
POP_CL <- data_CL_V_multi[ind_data_CL_V_multi,1]
```

---

```
POP_V <- data_CL_V_multi[ind_data_CL_V_multi,2]
```

The *dcat* function is used to extract an index (*ind\_data\_CL\_V\_multi*), which represents a row of the table *data\_CL\_V\_multi* (defined in the dataset). The probability of each row (*weight\_CL\_V*), when it is specified, is retrieved from the original table, otherwise, a vector with equal probabilities is created and passed to *dcat*. Finally, the values on the *ind\_data\_CL\_V\_multi*-th row are extracted from *data\_CL\_V\_multi* and assigned to the corresponding variable.

### ***ODE model management***

For an ODE model, *model\_BUGS.txt* contains the call to a function, which has the form:

```
function.model1(grid[], amt[], rate[], ii[], evid[], cmt[], addl[], ss[], theta[,])
```

where *time*, *amt*, *rate*, *ii*, *evid*, *cmt*, *addl* and *ss* are  $1 \times \max(N_t)$  vectors, and *theta* is a matrix. *function.model1* is the name of the function linked to a Pascal model library component (*ODEPascal.txt*) where the system of differential equations to be solved is specified. The interface between WinBUGS and the Pascal model library component is performed using WBDev, in particular the Pascal file named *ODEPascal\_Mod.txt*. An important design feature is that all the individual variables are recalculated inside the Pascal model library component; for this reason, all the fixed effect and random effect variables (together with the covariates) must be included in *theta*. The variable *theta* is constructed in a loop over *ind\_t2* (see below) as a matrix with a number of rows equal to  $\max(N_t)$ . This structure is needed because the Pascal model library component that solves the ODE resets the clock to zero at each of the post-zero user-specified times. This can be handled by passing the start times via the *theta* argument. For this reason, the *ind\_t2*-th row of *theta* contains the first grid point of the time interval *grid[ind\_t2-1]*, *grid[ind\_t2]*. Then, the following elements are included (in this order):

- *n\_cov\_cont* (only if continuous covariates are used)
- *max\_m\_<CovariateName>* (only if continuous covariates are used)
- *N\_t\_<CovariateName>* (only if continuous covariates are used)
- *n\_cov\_cat* (only if categorical covariates are used)
- *max\_m\_<CovariateName>* (only if categorical covariates are used)
- *N\_t\_<CovariateName>* (only if categorical covariates are used)
- *grid\_<CovariateName>* (only if continuous /categorical covariates are used)
- *<CovariateName>* (only if continuous/categorical covariates are used)
- Fixed effect and random effect variables
- Initial conditions for each derivative variable (always)
- Fraction (F) value for each derivative variable (set to 1 by default) (always)

- Tlag (TLAG) value for each derivative variable (set to 0 by default) (always)

For example, in UseCase1:

```

for (ind_t2 in 1:N_t[ind_subj]) {
  index_unipv[ind_subj,ind_t2] <- max(1,ind_t2-1)
  theta[ind_subj,ind_t2,1] <- grid[ind_subj,index_unipv[ind_subj,ind_t2]] #t0
  theta[ind_subj,ind_t2,2] <- n_cov_cont
  theta[ind_subj,ind_t2,3] <- max_m_logtWT
  theta[ind_subj,ind_t2,4] <- N_t_logtWT[ind_subj]
  for (i in 1 : max_m_logtWT){
    theta[ind_subj,ind_t2,i+4] <- grid_logtWT[ind_subj,i]
  }
  for (i in 1 : max_m_logtWT){
    theta[ind_subj,ind_t2,i+4+max_m_logtWT] <- logtWT[ind_subj,i]
  }
  theta[ind_subj,ind_t2,4+2*max_m_logtWT +1] <- ETA_CL[ind_subj]
  theta[ind_subj,ind_t2,4+2*max_m_logtWT +2] <- POP_CL
  theta[ind_subj,ind_t2,4+2*max_m_logtWT +3] <- BETA_CL_WT
  theta[ind_subj,ind_t2,4+2*max_m_logtWT +4] <- ETA_V[ind_subj]
  theta[ind_subj,ind_t2,4+2*max_m_logtWT +5] <- POP_V
  theta[ind_subj,ind_t2,4+2*max_m_logtWT +6] <- BETA_V_WT
  theta[ind_subj,ind_t2,4+2*max_m_logtWT +7] <- ETA_KA[ind_subj]
  theta[ind_subj,ind_t2,4+2*max_m_logtWT +8] <- POP_KA
  theta[ind_subj,ind_t2,4+2*max_m_logtWT +9] <- ETA_TLAG[ind_subj]
  theta[ind_subj,ind_t2,4+2*max_m_logtWT +10] <- POP_TLAG
  theta[ind_subj,ind_t2,4+2*max_m_logtWT+11] <- 0 #IC GUT
  theta[ind_subj,ind_t2,4+2*max_m_logtWT+12] <- 0 #IC CENTRAL
  theta[ind_subj,ind_t2,4+2*max_m_logtWT +13] <- 1 #F GUT
  theta[ind_subj,ind_t2,4+2*max_m_logtWT+14] <- 1 #F CENTRAL
  theta[ind_subj,ind_t2,4+2*max_m_logtWT+15] <- 0 #TLAG GUT
  theta[ind_subj,ind_t2,4+2*max_m_logtWT+16] <- 0 #TLAG CENTRAL

```

The output of *function.modell* is stored in the *der99wb\_unipv* variable, which is a matrix with a number of rows equal to the length of *grid* and a number of columns equal to the number of derivative variables. Each derivative variable is then retrieved and saved in specific variables in another loop over *ind\_q*, as (for UseCase1):

```

for (ind_q in 1:N_t[ind_subj]) {
  GUT[ind_subj,ind_q] <- der99wb_unipv[ind_subj,ind_q,1] + 0
  CENTRAL[ind_subj,ind_q] <- der99wb_unipv[ind_subj,ind_q,2] + 0
  CC[ind_subj,ind_q] <- (CENTRAL[ind_subj,ind_q] / V[ind_subj,ind_q])
}

```

The same structure is used when the drug input, distribution, and elimination is described with a modular approach (MDL Compartments, using PK macros). Modules are converted into the corresponding differential equations, setting appropriate values for F and TLAG when specified in the MDL structure.

In the Pascal model library component named *ODEPascal.txt*, the ODEs are specified. This element is originated from a template where the following parts are modified any time according to the model specified by the user:

- 
- The number of differential equations by editing the line *numEq*=\*.
  - The differential equations by specifying the model inside the *UserDerivatives* procedure. All values passed from WinBUGS via the theta vector are here imported (indexes in Pascal start with 0). Due to a design choice of the BUGSModelLibrary, the Pascal model library component that solves the ODE system sets all the initial conditions to 0. To allow the specification of initial conditions different from 0, the model is automatically generated in terms of difference from initial condition ( $z(t)=x(t)-x_0$ , where  $z$  is the new computed variable,  $x$  is the original state variable and  $x_0$  its initial condition). For example, in UseCase1:

```
VAR0 := DER99WB_UNIPV[0]+IC0;
VAR1 := DER99WB_UNIPV[1]+IC1;
dDER99WB_UNIPVdt[0] := -(piecewise_RATEIN_UNIPV));
dDER99WB_UNIPVdt[1] := (piecewise_RATEIN_UNIPV - ((CL_UNIPV *
VAR1) / V_UNIPV));
```

Then, in the WinBUGS code the initial conditions (IC0 and IC1) are eventually added to the resulting ODE solutions ( $x(t)=z(t)+x_0$ ) in the loop over *ind\_q*. For example, in UseCase1 (where both IC0 and IC1 are 0):

```
for (ind_q in 1:N_t[ind_subj]) {
GUT[ind_subj,ind_q] <- der99wb_unipv[ind_subj,ind_q,1] + 0
CENTRAL[ind_subj,ind_q] <- der99wb_unipv[ind_subj,ind_q,2] + 0
}
```

- Interpolation of continuous and categorical covariates over the grid of the ODE solver using the following procedures: *InterpMoreCov3* for continuous covariates and *InterpMoreCov3\_CAT* for categorical covariates.
- Variable name declarations in the section following *VAR* inside the *UserDerivatives* procedure.
- The parameters in the *InitModel* procedure by editing *m.nParameter* (total number of theta elements), *m.FIIndex* (index of the first F in theta), *m.tlag1Index* (index of first TLAG in theta), *m.nCmt* (number of compartments).

### **Conditional statements management**

Every individual variable defined via an IF-THEN-ELSE statement in MDL is calculated in *model\_BUGS.txt*, using the function named *function.piecewise\** (linked to the Pascal model library component named *FunctionPiecewise\*.txt*). This function takes as input all the covariates (already interpolated), fixed effect and random effect variables necessary to evaluate the conditional statement. The function call is located inside the loop over *ind\_q* if the computation of the IF-THEN-ELSE statement involves at least one derivative variable. For example, in UseCase1:

---

```

piecewise_RATEIN[ind_subj,ind_q]<-
function.piecewise3(VECTOR1[ind_subj,ind_q,])
VECTOR1[ind_subj,ind_q,1] <- GUT[ind_subj,ind_q]
VECTOR1[ind_subj,ind_q,2] <- grid[ind_subj,ind_q]
VECTOR1[ind_subj,ind_q,3] <- ETA_KA[ind_subj]
VECTOR1[ind_subj,ind_q,4] <- POP_KA
VECTOR1[ind_subj,ind_q,5] <- ETA_TLAG[ind_subj]
VECTOR1[ind_subj,ind_q,6] <- POP_TLAG

```

*FunctionPiecewise\*.txt* component is originated from a template where the Evaluate procedure is modified every time. This procedure calls *func.arguments[0][i].Value()* which returns the value of the i-th element of the function argument passed from WinBUGS via the *VECTOR\** vector, then the value returned by the conditional statements is stored in a variable called *value*.

A proper function named *function.covariate\** (linked to the corresponding Pascal model library component named *FunctionCovariate\*.txt*) is used similarly to map the arbitrary numeric codes defined in the dataset of a binary categorical variable into 0-1 values and to manage categorical variables (with more than two levels) when these are used in the definition of individual parameters via a structured model. In the latter case, ( $n^\circ$  of levels -1) dummy variables, which can assume only 0 and 1 values, are created to reproduce the different categories.

### **Multivariate observation model**

A specific model encoding in MDL is required when multiple observations with correlated errors are present. For example, in M4 (148):

```

magni2000_M4_mdI = mdlObj {
  OBSERVATION {
    Y1 : {additive = 1, eps = EPS1, prediction = ts_PRED, type is additiveError}
    Y2 : {additive = 1, eps = EPS2, prediction = F_PRED, type is additiveError}
    Y3 : {additive = 1, eps = EPS3, prediction = tl_PRED, type is additiveError}
    Y4 : {additive = 1, eps = EPS4, prediction = V_PRED, type is additiveError}
    EPS1=EPS[1]
    EPS2=EPS[2]
    EPS3=EPS[3]
    EPS4=EPS[4]
  }
  VARIABILITY_LEVELS {
    DV : {level = 1, type is observation}
  }
  RANDOM_VARIABLE_DEFINITION(level=DV) {
    EPS ~ MultivariateNormal2(mean=[0,0,0,0],
    precisionMatrix=invOMEGA_PAR)
  }
  #other code
}

```

The following requirements are mandatory:

- 
- In the *RANDOM\_VARIABLE\_DEFINITION* block the residual error (*EPS*) must be distributed as a *MultivariateNormal* with variability level set to *DV*;
  - In the *OBSERVATION* block, for each observation, the residual error model must be “additive” with the parameter *additive* set to 1 and the parameter *eps* set to a variable defined in the same block which is equal to the corresponding element of the *EPS* vector.

In WinBUGS, the MDL code reported above is translated into:

```

Y1_mean[ind_subj,ind_t,1] <- ts_PRED[ind_subj,ind_t]
Y1_mean[ind_subj,ind_t,2] <- F_PRED[ind_subj,ind_t]
Y1_mean[ind_subj,ind_t,3] <- tl_PRED[ind_subj,ind_t]
Y1_mean[ind_subj,ind_t,4] <- V_PRED[ind_subj,ind_t]
Y1[ind_subj,ind_t,1:4] ~
dmnorm(Y1_mean[ind_subj,ind_t,],invOMEGA_PAR[,])

```

### ***RunScript.txt structure***

It contains the commands to check the model syntax, load the data, compile the model, generate randomly (or provide) initial values for the initialized chain(s), update the chain(s) with the burn-in iterations, set the monitored nodes, update the chain(s) with the iterations that will be kept for the analysis, and, finally, save the samples values of all monitored nodes in specific CODA files. There is also the possibility to set the thin (by default it is 1) and the seed (by default is not set). The developed WinBUGS plugin supports the generation of only one Markov chain. By default, the monitored variables are all the stochastic nodes having a prior distribution, the individual predictions (*Y\_mean*), and the individual residuals (*Y\_res*).

For example, in UseCase1:

```

display("log")
#model: UseCase1
#check if the model is syntactically correct
check('model_BUGS.txt')
#load data
data('data_BUGS.txt')
#load seed
#set.seed(%s)
#compile the model
compile(1)
#load initial estimates (by default is commented)
#inits(1,'inits_BUGS.txt')
#inits has as first argument the number of the chain (default=1) and the
filename with the initial estimates
#generate random initial estimates
gen.inits()
thin.updater(1)
thin.samples(1)
#burn-in (commented if simulation)
update(100)
set(POP_CL)
set(POP_V)

```

```

set(POP_KA)
set(POP_TLAG)
set(PPV_CL)
set(PPV_V)
set(PPV_KA)
set(RUV_PROP)
set(RUV_ADD)
set(Y_res)
set(Y_mean)
update(1000)
#save results
coda(*,'output')
quit('yes')

```

## Supported operators

The unary and binary operators supported by the converter in the WinBUGS plugin are herein described. The PharmML operator name is reported in the first column, the corresponding translation into WinBUGS and Pascal code are reported in the second and third columns, respectively. Only the operators supported both in BUGS and in Pascal languages are allowed.

**Table A.S1:** Supported operators.

PharmML	WinBUGS	Component Pascal
<b>Binary operations</b>		
divide(a,b)	a/b	a/b
power(a,b)	pow(a,b)	Math.Power(a,b)
min(a,b)	min(a,b)	MIN(a,b)
max(a,b)	max(a,b)	MAX(a,b)
minus(a,b)	a-b	a-b
plus(a,b)	a+b	a+b
times(a,b)	a*b	a*b
root(a,b)	pow(a,1/b)	Math.Power(a,1/b)
logx(a,b)	log(a)/log(b)	Math.Ln(a)/Math.Ln(b)
<b>Unary operations</b>		
abs(a)	abs(a)	ABS(a)
log(a)	log(a)	Math.Ln(a)
logit(a)	logit(a)	Math.Ln(a)-Math.Ln(1-a)
sqrt(a)	sqrt(a)	Math.Sqrt(a)
exp(a)	exp(a)	Math.Exp(a)
factln(a)	logfact(a)	MathFunc.LogFactorial(a)
floor(a)	trunc(a)	Math.Floor(a)
gammln(a)	loggam(a)	MathFunc.LogGammaFunc(a)
minus(a)	-(a)	-(a)

PharmML	WinBUGS	Component Pascal
normcdf(a)	phi(a)	MathFunc.Phi(a)
factorial(a)	exp(logfact(a))	Math.Exp(MathFunc.LogFactorial(a))
ceiling(a)	trunc(a)+1	Math.Ceiling(a)
logistic(a)	1/(1+exp(-a))	1/(1+Math.Exp(-a))
sin(a)	sin(a)	Math.Sin(a)
cos(a)	cos(a)	Math.Cos(a)
tan(a)	sin(a)/cos(a)	Math.Tan(a)
sec(a)	1/cos(a)	1/Math.Cos(a)
csc(a)	1/sin(a)	1/Math.Sin(a)
cot(a)	cos(a)/sin(a)	Math.Cos(a)/Math.Sin(a)
sinh(a)	(exp(a)-exp(-a))/2	Math.Sinh(a)
cosh(a)	(exp(a)+exp(-a))/2	Math.Cosh(a)
tanh(a)	(exp(a)-exp(-a))/(exp(a)+exp(-a))	Math.Tanh(a)
coth(a)	(exp(a)+exp(-a))/(exp(a)-exp(-a))	(Math.Exp(a)+Math.Exp(-a))/(Math.Exp(a)-Math.Exp(-a))
sech(a)	2/(exp(a)+exp(-a))	2/(Math.Exp(a)+Math.Exp(-a))
csch(a)	2/(exp(a)-exp(-a))	2/(Math.Exp(a)-Math.Exp(-a))
arcsinh(a)	log(a+sqrt(pow(a,2)+1))	Math.ArcSinh(a)
arccosh(a)	log(a+sqrt(pow(a,2)-1))	Math.ArcCosh(a)
arctanh(a)	1/2*log((1+a)/(1-a))	Math.ArcTanh(a)
arccoth(a)	1/2*log((a+1)/(a-1))	1/2*Math.Log((a+1)/(a-1))
arcsech(a)	Not supported	Not supported
arccsch(a)	Not supported	Not supported
arcsin(a)	Not supported	Math.ArcSin(a)
arccos(a)	Not supported	Math.ArcCos(a)
arctan(a)	Not supported	Math.ArcTan(a)
arcsec(a)	Not supported	Math.ArcCos(1/a)
arccsc(a)	Not supported	Math.ArcSin(1/a)
arccot(a)	Not supported	Math.ArcTan(1/a)
probit(a)	Not supported	Not supported

## Supported probability distributions

The probability distributions supported by the WinBUGS plugin are herein described. All the supported distributions and arguments are given according to the ProbOnto v.2.0 knowledge-base (150), which is used by both MDL and PharmML. Distribution names are reported in the first column, their parameters in the second column, and the corresponding WinBUGS translation in the third column (in some cases transformations on parameters, here specified, are required).

**Table A.S2:** Supported probability distributions.

ProbOnto	Parameters	WinBUGS code
BETA1		
	alpha	$x \sim \text{dbeta}(\text{alpha}, \text{beta})$
	beta	
EXPONENTIAL1		
	rate	$x \sim \text{dexp}(\text{rate})$
EXPONENTIAL2		
	mean	$x \sim \text{dexp}(\text{rate})$
		$\text{rate} <- 1/\text{mean}$
GAMMA1		
	shape	$x \sim \text{dgamma}(\text{shape}, \text{rate})$
	scale	$\text{rate} <- 1/\text{scale}$
GAMMA2		
	shape	$x \sim \text{dgamma}(\text{shape}, \text{rate})$
	rate	
LOGNORMAL1		
	meanLog	$x \sim \text{dlnorm}(\text{meanLog}, \text{precision})$
	stdevLog	$\text{precision} <- 1/\text{pow}(\text{stdevLog}, 2)$
LOGNORMAL2		
	meanLog	$x \sim \text{dlnorm}(\text{meanLog}, \text{precision})$
	varLog	$\text{precision} <- 1/\text{varLog}$
LOGNORMAL3		
	median	$x \sim \text{dlnorm}(\text{meanLog}, \text{precision})$
	stdevLog	$\text{meanLog} <- \log(\text{median})$
		$\text{precision} <- 1/\text{pow}(\text{stdevLog}, 2)$
LOGNORMAL4		
	median	$x \sim \text{dlnorm}(\text{meanLog}, \text{precision})$
	coefVar	$\text{meanLog} <- \log(\text{median})$
		$\text{precision} <- 1/(\log(\text{pow}(\text{coefVar}, 2)+1))$
LOGNORMAL5		
	meanLog	$x \sim \text{dlnorm}(\text{meanLog}, \text{precision})$

ProbOnto	Parameters	WinBUGS code
	precision	
LOGNORMAL6		
	median	$x \sim \text{dlnorm}(\text{meanLog}, \text{precision})$
	geomStdev	$\text{meanLog} \leftarrow -\log(\text{median})$
		$\text{precision} \leftarrow 1/\text{pow}(\log(\text{geomStdev}), 2)$
NORMAL1		
	mean	$x \sim \text{dnorm}(\text{mean}, \text{precision})$
	stdev	$\text{precision} \leftarrow 1/\text{pow}(\text{stdev}, 2)$
NORMAL2		
	mean	$x \sim \text{dnorm}(\text{mean}, \text{precision})$
	variance	$\text{precision} \leftarrow 1/\text{variance}$
NORMAL3		
	mean	$x \sim \text{dnorm}(\text{mean}, \text{precision})$
	precision	
STUDENTT1		
	degreesOfFreedom	$x \sim \text{dt}(\text{mean}, \text{scale}, \text{degreesOfFreedom})$
		$\text{mean} \leftarrow 0$
		$\text{scale} \leftarrow 1$
STUDENTT2		
	mean	$x \sim \text{dt}(\text{mean}, \text{scale}, \text{degreesOfFreedom})$
	scale	
	degreesOfFreedom	
UNIFORM1		
	minimum	$x \sim \text{dunif}(\text{minimum}, \text{maximum})$
	maximum	
WEIBULL1		
	scale	$x \sim \text{dweib}(\text{lambda}, \text{shape})$
	shape	$\text{lambda} \leftarrow \text{pow}(\text{scale}, -1/\text{shape})$
WEIBULL2		
	lambda	$x \sim \text{dweib}(\text{lambda}, \text{shape})$
	shape	
INVERSEGAMMA1		
	shape	$y \sim \text{dgamma}(\text{shape}, \text{rate})$
	scale	$\text{rate} \leftarrow 1/\text{scale}$
		$x \leftarrow 1/y$
<b>Continuous Multivariate Distributions</b>		
MULTIVARIATE NORMAL1		
	mean	$x[1:\text{dim}] \sim \text{dmnorm}(\text{mean}[], \text{precisionMatrix}[,])$
	covarianceMatrix	$\text{precisionMatrix}[1:\text{dim}, 1:\text{dim}] \leftarrow \text{inverse}(\text{covarianceMatrix}[,])$

ProbOnto	Parameters	WinBUGS code
MULTIVARIATE NORMAL2		
	mean	$x[1:dim] \sim dnmnorm(mean[], precisionMatrix[,])$
	precisionMatrix	
WISHART1		
	scaleMatrix	$x[1:dim, 1:dim] \sim dwish(inverseScaleMatrix[,], degreesOfFreedom)$
	degreesOfFreedom	$inverseScaleMatrix[1:dim, 1:dim] \leftarrow inverse(scaleMatrix[,])$
WISHART2		
	inverseScaleMatrix	$x[1:dim, 1:dim] \sim dwish(inverseScaleMatrix[,], degreesOfFreedom)$
	degreesOfFreedom	
INVERSE WISHART1		
	scaleMatrix	$inverseX[1:dim, 1:dim] \sim dwish(inverseScaleMatrix[,], degreesOfFreedom)$
	degreesOfFreedom	$inverseScaleMatrix[1:dim, 1:dim] \leftarrow inverse(scaleMatrix[,])$
		$x[1:dim, 1:dim] \leftarrow inverse(inverseX[,])$
MULTIVARIATE STUDENTT1		
	mean	$x[1:dim] \sim dmt(mean[], precisionMatrix[,], degreesOfFreedom)$
	covarianceMatrix	$precisionMatrix[1:dim, 1:dim] \leftarrow inverse(covarianceMatrix[,])$
	degreesOfFreedom	
MULTIVARIATE STUDENTT2		
	mean	$x[1:dim] \sim dmt(mean[], precisionMatrix[,], degreesOfFreedom)$
	precisionMatrix	
	degreesOfFreedom	

---

## List of Abbreviations

---

<b>DEEP</b>	DEferiprone Evaluation in Paediatrics
<b>DDMoRe</b>	Drug Disease Model Resources
<b>NME</b>	New Molecular Entities
<b>M&amp;S</b>	Modeling & Simulation
<b>MID3</b>	Model-Informed Drug Discovery & Development
<b>EMA</b>	European Medicines Agency
<b>FDA</b>	Food and Drug Administration
<b>ADME</b>	Absorption, Distribution, Metabolism, and Excretion
<b>PK</b>	Pharmacokinetic(s)
<b>PD</b>	Pharmacodynamic(s)
<b>NLME</b>	Non-Linear Mixed Effect
<b>IIV</b>	Inter-Individual Variability
<b>FIM</b>	Fisher Information Matrix
<b>CTS</b>	Clinical Trial Simulation(s)
<b>IMI</b>	Innovative Medicines Initiative
<b>MDL</b>	Modeling Description Language
<b>PharmML</b>	Pharmacometrics Markup Language
<b>SO</b>	Standard Output
<b>IOF</b>	Interoperability Framework
<b>BSA</b>	Body Surface Area

---

<b>EU</b>	European Union
<b>EEA</b>	European Economic Area
<b>UK</b>	United Kingdom
<b>MDS</b>	Myelodysplastic Syndromes
<b>AA</b>	Aplastic Anemia
<b>SCD</b>	Sickle-Cell Disease
<b>UGT</b>	UDP-glucuronosyltransferase
<b>CYP</b>	Cytochrome p450
<b>GOF</b>	Goodness-Of-Fit
<b>CWRES</b>	Conditional Weighted Residual
<b>VPC</b>	Visual Predictive Check
<b>NPDE</b>	Normalized Predictive Distribution Error
<b>PMA</b>	Post Menstrual Age
<b>DF</b>	Degrees of Freedom
<b>RSE</b>	Relative Standard Error
<b>SE</b>	Standard Error
<b>CV</b>	Coefficient of Variation
<b>AUC</b>	Area Under Curve
<b>C<sub>max</sub></b>	Maximum Concentration
<b>CI</b>	Confidence Interval
<b>PI</b>	Prediction Interval
<b>Hb</b>	Hemoglobin
<b>RBC</b>	Red Blood Cells
<b>F<sub>pn</sub></b>	Ferroportin
<b>Fe-Tf</b>	Iron (Fe) bounded to transferrin (Tf)
<b>L1</b>	Deferiprone

---

<b>XJ</b>	Deferasirox
<b>C<sub>ss</sub><sup>AV</sup></b>	Steady-State Average Concentration
<b>ISV</b>	Inter-Study Variability
<b>LIC</b>	Liver Iron Content
<b>EBE</b>	Empirical Bayes Estimates
<b>MAP</b>	Maximum A Posteriori
<b>IU</b>	International Unit
<b>TP</b>	True Positive
<b>TN</b>	True Negative
<b>FP</b>	False Positive
<b>FN</b>	False Negative
<b>PPV</b>	Positive-Predictive Value
<b>NPV</b>	Negative-Predictive Value
<b>ANOVA</b>	Analysis Of Variance
<b>MMRM</b>	Mixed Model for Repeated Measures
<b>MRI</b>	Magnetic Resonance Image
<b>CDF</b>	Cumulative Distribution Function
<b>MCMC</b>	Monte Carlo Markov Chain
<b>WBDev</b>	WinBUGS Development Interface
<b>ODE</b>	Ordinary Differential Equation
<b>MDL-IDE</b>	MDL-Integrated Development Environment
<b>IVGTT</b>	Intravenous Glucose Tolerance Test
<b>MM</b>	Minimal Model
<b>CP</b>	C-peptide
<b>ISR</b>	Insulin Secretion Rate
<b>MLE</b>	Maximum Likelihood Estimation

---

**SPI**

Service Provider Interface

---

## References

---

- [1] Pammolli F, Magazzini L, Riccaboni M. The productivity crisis in pharmaceutical R&D. *Nat Rev Drug Discov.* 2011 Jun;10(6):428–38.
- [2] Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat Rev Drug Discov.* 2012 Mar 1;11(3):191–200.
- [3] Cohen FJ. Macro trends in pharmaceutical innovation. *Discov Med.* 2005 Apr;5(26):153–8.
- [4] Leil TA, Bertz R. Quantitative Systems Pharmacology can reduce attrition and improve productivity in pharmaceutical research and development. *Front Pharmacol.* 2014 Nov 10;5.
- [5] Knight-Schrijver VR, Chelliah V, Cucurull-Sanchez L, Le Novère N. The promises of quantitative systems pharmacology modelling for drug development. *Comput Struct Biotechnol J.* 2016;14:363–70.
- [6] Marshall SF, Burghaus R, Cosson V, Cheung SYA, Chenel M, DellaPasqua O, et al. Good practices in model-informed drug discovery and development: practice, application, and documentation. *CPT Pharmacomet Syst Pharmacol.* 2016;5(3):93–122.
- [7] Manolis E, Rohou S, Hemmings R, Salmonson T, Karlsson M, Milligan PA. The role of modeling and simulation in development and registration of medicinal products: output from the EFPIA/EMA modeling and simulation workshop. *CPT Pharmacomet Syst Pharmacol.* 2013 Feb;2(2):e31.
- [8] Gobburu JV, Lesko LJ. Quantitative disease, drug, and trial models. *Annu Rev Pharmacol Toxicol.* 2009;49:291–301.

- 
- [9] US Food and Drug Administration. FDA's plan to advance regulatory science. [Internet]. 2011. Available from: <https://www.fda.gov/downloads/scienceresearch/specialtopics/regulatoryscience/ucm268225.pdf>
- [10] US Food and Drug Administration. Innovation or stagnation: challenge and opportunity on the critical path to new medical products. 2004.
- [11] Standing JF. Understanding and applying pharmacometric modelling and simulation in clinical practice and research. *Br J Clin Pharmacol*. 2017 Feb;83(2):247–54.
- [12] Murphy MF. Rare diseases: meeting the unique challenges of orphan drug development [Internet]. Available from: <http://www.appliedclinicaltrials.com/rare-diseases-meeting-unique-challenges-orphan-drug-development?pageID=2>
- [13] Musuamba FT, Manolis E, Holford N, Cheung S, Friberg LE, Ogungbenro K, et al. Advanced methods for dose and regimen finding during drug development: summary of the EMA/EFPIA workshop on dose finding (London 4-5 December 2014). *CPT Pharmacomet Syst Pharmacol*. 2017 Jul;6(7):418–29.
- [14] EFPIA MID3 Workgroup. Model Informed Drug Discovery and Development (MID3) Good Practice: use of prior knowledge and setting up assumptions [Internet]. 2015. Available from: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Presentation/2016/05/WC500207582.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2016/05/WC500207582.pdf)
- [15] DEEP – DEFERIPRONE EVALUATION IN PAEDIATRICS FP7-Health- F4-2010-261483.
- [16] European Medicines Agency. Committee for Medicinal Products for Human Use (CHMP). Guideline on clinical trials in small populations. Doc. Ref. CHMP/EWP/83561/2005. [Internet]. 2006. Available from: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/09/WC500003615.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf)

- 
- [17] US Food and Drug Administration. Report: complex issues in developing drugs and biological products for rare diseases and accelerating the development of therapies for pediatric rare diseases [Internet]. 2014. Available from: <https://www.fda.gov/downloads/RegulatoryInformation/LawsEnforcedbyFDA/SignificantAmendmentstotheFDCAAct/FDASIA/UCM404104.pdf>
- [18] Bellanti F, Della Pasqua O. Modelling and simulation as research tools in paediatric drug development. *Eur J Clin Pharmacol*. 2011;67(1):75–86.
- [19] De Cock RF, Piana C, Krekels EH, Danhof M, Allegaert K, Knibbe CA. The role of population PK–PD modelling in paediatric clinical research. *Eur J Clin Pharmacol*. 2011;67(1):5–16.
- [20] Nyberg J, Ueckert S, Strömberg EA, Hennig S, Karlsson MO, Hooker AC. PopED: an extended, parallelized, nonlinear mixed effects models optimal design tool. *Comput Methods Programs Biomed*. 2012;108(2):789–805.
- [21] Foracchia M, Hooker AC, Vicini P, Ruggeri A. POPED, a software for optimal experiment design in population kinetics. *Comput Methods Programs Biomed*. 2004;74(1):29–46.
- [22] Gamalo-Siebers M, Savic J, Basu C, Zhao X, Gopalakrishnan M, Gao A, et al. Statistical modeling for Bayesian extrapolation of adult clinical trial information in pediatric drug evaluation. *Pharm Stat* [Internet]. 2017; Available from: <http://onlinelibrary.wiley.com/doi/10.1002/pst.1807/full>
- [23] D’Argenio DZ. Optimal sampling times for pharmacokinetic experiments. *J Pharmacokinet Pharmacodyn*. 1981;9(6):739–756.
- [24] D’Argenio DZ. Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments. *Math Biosci*. 1990;99(1):105–118.
- [25] Bellanti F, Del Vecchio GC, Putti MC, Maggio A, Filosa A, Cosmi C, et al. Population pharmacokinetics and dosing recommendations for the use of deferiprone in children younger than 6 years. *Br J Clin Pharmacol*. 2017;83(3):593–602.

- 
- [26] Petit C, Jullien V, Samson A, Guedj J, Kiechel J-R, Zohar S, et al. Designing a pediatric study for an antimalarial drug by using information from adults. *Antimicrob Agents Chemother.* 2016;60(3):1481–1491.
- [27] Berman RE, Field MJ, others. The ethical conduct of clinical research involving children [Internet]. National Academies Press; 2004 [cited 2017 Sep 5]. Available from: <https://books.google.it/books?hl=it&lr=&id=El-bAgAAQBAJ&oi=fnd&pg=PT13&dq=The+ethical+conduct+of+clinical+research+involving+children&ots=vrzHKpRtbJ&sig=iimEYG-X4oW0UDGk8GJli-L63kc>
- [28] 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 Pharmacomet Syst Pharmacol.* 2016;5(6):305–312.
- [29] Smania G, Baiardi P, Ceci A, Cella M, Magni P. Model-based assessment of alternative study designs in pediatric trials. Part II: Bayesian approaches. *CPT Pharmacomet Syst Pharmacol.* 2016;5(8):402–410.
- [30] Chirnomas D, Smith AL, Braunstein J, Finkelstein Y, Pereira L, Bergmann AK, et al. Deferasirox pharmacokinetics in patients with adequate versus inadequate response. *Blood.* 2009;114(19):4009–4013.
- [31] Miyazawa K, Ohyashiki K, Urabe A, Hata T, Nakao S, Ozawa K, et al. A safety, pharmacokinetic and pharmacodynamic investigation of deferasirox (Exjade®, ICL670) in patients with transfusion-dependent anemias and iron-overload: a Phase I study in Japan. *Int J Hematol.* 2008;88(1):73.
- [32] Sechaud R, Robeva A, Belleli R, Balez S. Absolute oral bioavailability and disposition of deferasirox in healthy human subjects. *J Clin Pharmacol.* 2008;48(8):919–925.
- [33] Waldmeier F, Bruin GJ, Glaenzel U, Hazell K, Sechaud R, Warrington S, et al. Pharmacokinetics, metabolism, and disposition of deferasirox in  $\beta$ -thalassemic patients with transfusion-dependent iron overload who are at pharmacokinetic steady state. *Drug Metab Dispos.* 2010;38(5):808–816.

- 
- [34] Galanello R, Piga A, Cappellini MD, Forni GL, Zappu A, Origa R, et al. Effect of food, type of food, and time of food intake on deferasirox bioavailability: recommendations for an optimal deferasirox administration regimen. *J Clin Pharmacol*. 2008;48(4):428–435.
- [35] Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 2006;34(suppl\_1):D668–D672.
- [36] Walpole SC, Prieto-Merino D, Edwards P, Cleland J, Stevens G, Roberts I. The weight of nations: an estimation of adult human biomass. *BMC Public Health*. 2012;12(1):439.
- [37] Average Height to Weight Chart - Babies to Teenagers [Internet]. Disabled World. [cited 2017 Sep 5]. Available from: <https://www.disabled-world.com/artman/publish/height-weight-teens.shtml>
- [38] Figg W, McLeod HL. Handbook of anticancer pharmacokinetics and pharmacodynamics [Internet]. Springer Science & Business Media; 2004. Available from: <https://books.google.it/books?hl=it&lr=&id=VIT5BwAAQBAJ&oi=fnd&pg=PR5&dq=Handbook+of+anticancer+pharmacokinetics+and+pharmacodynamics&ots=NlcclLp-a9&sig=9gvBx5-RZ99rylmYa3diwDPGn3U>
- [39] Miyagi SJ, Milne AM, Coughtrie MW, Collier AC. Neonatal development of hepatic UGT1A9: implications of pediatric pharmacokinetics. *Drug Metab Dispos*. 2012;40(7):1321–1327.
- [40] Strassburg CP, Strassburg A, Kneip S, Barut A, Tukey RH, Rodeck B, et al. Developmental aspects of human hepatic drug glucuronidation in young children and adults. *Gut*. 2002;50(2):259–265.
- [41] Tayman C, Rayyan M, Allegaert K. Neonatal pharmacology: extensive interindividual variability despite limited size. *J Pediatr Pharmacol Ther*. 2011;16(3):170–184.
- [42] Walton K, Dorne JL, Renwick AG. Uncertainty factors for chemical risk assessment: interspecies differences in glucuronidation. *Food Chem Toxicol*. 2001;39(12):1175–1190.

- 
- [43] Bruin GJ, Faller T, Wiegand H, Schweitzer A, Nick H, Schneider J, et al. Pharmacokinetics, distribution, metabolism, and excretion of deferasirox and its iron complex in rats. *Drug Metab Dispos.* 2008;36(12):2523–2538.
- [44] Sumpter AL, Holford NH. Predicting weight using postmenstrual age—neonates to adults. *Pediatr Anesth.* 2011;21(3):309–315.
- [45] Gisleskog PO, Karlsson MO, Beal SL. Use of prior information to stabilize a population data analysis. *J Pharmacokinet Pharmacodyn.* 2002;29(5):473–505.
- [46] US Food and Drug Administration. Center for Drug Evaluation and Research (CDER). Draft Guidance: Drug interaction studies: study design, data analysis, implications for dosing, and labeling recommendations. [Internet]. 2012. Available from: <https://www.fda.gov/downloads/drugs/guidances/ucm292362.pdf>
- [47] European Medicines Agency. Reflection paper on extrapolation of efficacy and safety in paediatric medicine development. Doc. Ref. EMA/199678/2016. [Internet]. 2016. Available from: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Regulatory\\_and\\_procedural\\_guideline/2016/04/WC500204187.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2016/04/WC500204187.pdf)
- [48] Dunne J, Rodriguez WJ, Murphy MD, Beasley BN, Burckart GJ, Filie JD, et al. Extrapolation of adult data and other data in pediatric drug-development programs. *Pediatrics.* 2011;pedS–2010.
- [49] European Medicines Agency (EMA). Committee for Medicinal Products for Human Use (CHMP). Guideline on the role of pharmacokinetics in the development of medicinal products in the paediatric population. Doc. Ref. EMA/CHMP/EWP/147013/2004 Corrigendum [Internet]. 2006. Available from: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/09/WC500003066.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003066.pdf)
- [50] Manolis E, Pons G. Proposals for model-based paediatric medicinal development within the current European Union regulatory framework. *Br J Clin Pharmacol.* 2009;68(4):493–501.
- [51] Anderson BJ, Holford NH. Mechanistic basis of using body size and maturation to predict clearance in humans. *Drug Metab Pharmacokinet.* 2009;24(1):25–36.

- 
- [52] Bouwmeester NJ, Anderson BJ, Tibboel D, Holford NH. Developmental pharmacokinetics of morphine and its metabolites in neonates, infants and young children. *Br J Anaesth*. 2004;92(2):208–217.
- [53] Benjamin DK, Smith PB, Jadhav P, Gobburu JV, Murphy MD, Hasselblad V, et al. Pediatric antihypertensive trial failures. *Hypertension*. 2008;51(4):834–840.
- [54] Momper JD, Karesh A, Green DJ, Hirsch M, Khurana M, Lee J, et al. Drug development for pediatric neurogenic bladder dysfunction: dosing, endpoints, and study design. *J Clin Pharmacol*. 2014;54(11):1239–1246.
- [55] Steinbicker AU, Muckenthaler MU. Out of balance—systemic iron homeostasis in iron-related disorders. *Nutrients*. 2013;5(8):3034–3061.
- [56] Andrews NC. Iron homeostasis: insights from genetics and animal models. *Nat Rev Genet*. 2000;1(3):208.
- [57] Ganz T. Systemic iron homeostasis. *Physiol Rev*. 2013;93(4):1721–1741.
- [58] Knovich MA, Storey JA, Coffman LG, Torti SV, Torti FM. Ferritin for the clinician. *Blood Rev*. 2009;23(3):95–104.
- [59] Millonig G, Muckenthaler MU, Mueller S. Hyperferritinaemia-cataract syndrome: worldwide mutations and phenotype of an increasingly diagnosed genetic disorder. *Hum Genomics*. 2010;4(4):250.
- [60] Kell DB, Pretorius E. Serum ferritin is an important inflammatory disease marker, as it is mainly a leakage product from damaged cells. *Metallomics*. 2014;6(4):748–773.
- [61] Porter JB. Practical management of iron overload. *Br J Haematol*. 2001;115(2):239–252.
- [62] Porter JB. Monitoring and treatment of iron overload: state of the art and new approaches. In: *Seminars in hematology* [Internet]. Elsevier; 2005. p. S14–S18. Available from: [http://www.seminhematol.org/article/S0037-1963\(05\)00005-3/abstract](http://www.seminhematol.org/article/S0037-1963(05)00005-3/abstract)
- [63] Porter JB, Huehns ER. The toxic effects of desferrioxamine. *Baillieres Clin Haematol*. 1989;2(2):459–474.

- 
- [64] Brittenham GM. Iron-chelating therapy for transfusional iron overload. *N Engl J Med*. 2011;364(2):146–156.
- [65] Theil EC. Mining ferritin iron: 2 pathways. *Blood*. 2009 Nov 12;114(20):4325–6.
- [66] Anderson LJ, Wonke B, Prescott E, Holden S, Walker JM, Pennell DJ. Comparison of effects of oral deferiprone and subcutaneous desferrioxamine on myocardial iron concentrations and ventricular function in beta-thalassaemia. *The Lancet*. 2002;360(9332):516–520.
- [67] Ho PJ, Tay L, Lindeman R, Catley L, Bowden DK. Australian guidelines for the assessment of iron overload and iron chelation in transfusion-dependent thalassaemia major, sickle cell disease and other congenital anaemias. *Intern Med J*. 2011;41(7):516–524.
- [68] Angelucci E, Barosi G, Camaschella C, Cappellini MD, Cazzola M, Galanello R, et al. Italian Society of Hematology practice guidelines for the management of iron overload in thalassemia major and related disorders. *haematologica*. 2008;93(5):741–752.
- [69] Porter JB, Elalfy M, Taher A, Aydinok Y, Lee S-H, Sutcharitchan P, et al. Limitations of serum ferritin to predict liver iron concentration responses to deferasirox therapy in patients with transfusion-dependent thalassaemia. *Eur J Haematol*. 2017 Mar 1;98(3):280–8.
- [70] Cappellini MD, Cohen A, Porter J, Taher A, Viprakasit V. Guidelines for the management of transfusion dependent thalassaemia (TDT) [Internet]. Thalassaemia International Federation Nicosia (CY); 2014. Available from: [https://www.researchgate.net/profile/Robert\\_Yamashita/publication/267451177\\_Psychological\\_Support/links/544fcea0cf24e8f7374a816.pdf](https://www.researchgate.net/profile/Robert_Yamashita/publication/267451177_Psychological_Support/links/544fcea0cf24e8f7374a816.pdf)
- [71] George E, Wong HB, George R, Ariffin WA. Serum ferritin concentrations intransfusion. *Singapore Med J*. 1994;35:62–64.
- [72] Aydinok Y, Ulger Z, Nart D, Terzi A, Cetiner N, Ellis G, et al. A randomized controlled 1-year study of daily deferiprone plus twice weekly desferrioxamine compared with daily deferiprone monotherapy in patients with thalassemia major. *Haematologica*. 2007;92(12):1599–1606.

- 
- [73] Ceci A, Baiardi P, Felisi M, Cappellini MD, Carnelli V, De Sanctis V, et al. The safety and effectiveness of deferiprone in a large-scale, 3-year study in Italian patients. *Br J Haematol.* 2002;118(1):330–336.
- [74] Choudhry VP, Pati HP, Saxena A, Malaviya AN. Deferiprone, efficacy and safety. *Indian J Pediatr.* 2004;71(3):213–216.
- [75] Hoffbrand AV, AL-Refaie F, Davis B, Siritanakatkul N, Jackson BF, Cochrane J, et al. Long-term trial of deferiprone in 51 transfusion-dependent iron overloaded patients. *Blood.* 1998;91(1):295–300.
- [76] Pennell DJ, Berdoukas V, Karagiorga M, Ladis V, Piga A, Aessopos A, et al. Randomized controlled trial of deferiprone or deferoxamine in beta-thalassemia major patients with asymptomatic myocardial siderosis. *Blood.* 2006;107(9):3738–3744.
- [77] Cappellini MD, Cohen A, Piga A, Bejaoui M, Perrotta S, Agaoglu L, et al. A phase 3 study of deferasirox (ICL670), a once-daily oral iron chelator, in patients with  $\beta$ -thalassemia. *Blood.* 2006;107(9):3455–3462.
- [78] Cappellini MD, Porter J, El-Beshlawy A, Li C-K, Seymour JF, Elalfy M, et al. Tailoring iron chelation by iron intake and serum ferritin: the prospective EPIC study of deferasirox in 1744 patients with transfusion-dependent anemias. *haematologica.* 2010;95(4):557–566.
- [79] Galanello R, Piga A, Forni GL, Bertrand Y, Foschini ML, Bordone E, et al. Phase II clinical evaluation of deferasirox, a once-daily oral chelating agent, in pediatric patients with beta-thalassemia major. *Haematologica.* 2006;91(10):1343–1351.
- [80] Piga A, Galanello R, Forni GL, Cappellini MD, Origa R, Zappu A, et al. Randomized phase II trial of deferasirox (Exjade, ICL670), a once-daily, orally-administered iron chelator, in comparison to deferoxamine in thalassemia patients with transfusional iron overload. *haematologica.* 2006;91(7):873–880.
- [81] Maggio A, D'Amico G, Morabito A, Capra M, Ciaccio C, Cianciulli P, et al. Deferiprone versus deferoxamine in patients with thalassemia major: a randomized clinical trial. *Blood Cells Mol Dis.* 2002;28(2):196–208.

- 
- [82] Olivieri NF, Nathan DG, MacMillan JH, Wayne AS, Liu PP, McGee A, et al. Survival in medically treated patients with homozygous  $\beta$ -thalassemia. *N Engl J Med.* 1994;331(9):574–578.
- [83] Pootrakul P, Sirankapracha P, Sankote J, Kachintorn U, Maungsub W, Sriphen K, et al. Clinical trial of deferiprone iron chelation therapy in  $\beta$ -thalassaemia/haemoglobin E patients in Thailand. *Br J Haematol.* 2003;122(2):305–310.
- [84] Olivieri NF, Brittenham GM, Matsui D, Berkovitch M, Blendis LM, Cameron RG, et al. Iron-chelation therapy with oral deferiprone in patients with thalassemia major. *N Engl J Med.* 1995;332(14):918–922.
- [85] Bellanti F, Danhof M, Della Pasqua O. Population pharmacokinetics of deferiprone in healthy subjects. *Br J Clin Pharmacol.* 2014;78(6):1397–1406.
- [86] Aydinok Y, Kattamis A, Viprakasit V. Current approach to iron chelation in children. *Br J Haematol.* 2014 Jun 1;165(6):745–55.
- [87] Cohen AR, Glimm E, Porter JB. Effect of transfusional iron intake on response to chelation therapy in beta-thalassemia major. *Blood.* 2008 Jan 15;111(2):583–7.
- [88] Musallam KM, Angastiniotis M, Eleftheriou A, Porter JB. Cross-Talk between Available Guidelines for the Management of Patients with Beta-Thalassemia Major. *Acta Haematol.* 2013;130(2):64–73.
- [89] Sheiner LB, Rosenberg B, Melmon KL. Modelling of individual pharmacokinetics for computer-aided drug dosage. *Comput Biomed Res Int J.* 1972 Oct;5(5):411–59.
- [90] Worwoon M, Cragg SJ, Jacobs A, McLaren C, Rickeits C, Economidou J. Binding of Serum Ferritin to Concanavalin A: Patients with Hornozygous  $\beta$  Thalassaemia and Transfusional Iron Overload. *Br J Haematol.* 1980;46(3):409–416.
- [91] Letsky EA, Miller F, Worwood M, Flynn DM. Serum ferritin in children with thalassaemia regularly transfused. *J Clin Pathol.* 1974;27(8):652–655.

- 
- [92] Rose C, Ernst O, Hecquet B, Maboudou P, Renom P, Noel MP, et al. Quantification by magnetic resonance imaging and liver consequences of post-transfusional iron overload alone in long-term survivors after allogeneic hematopoietic stem cell transplantation. *Haematologica*. 2007 Jun 1;92(6):850–3.
- [93] De Domenico I, Ward DM, Kaplan J. Specific iron chelators determine the route of ferritin degradation. *Blood*. 2009;114(20):4546–4551.
- [94] Zong X-N, Li H. Construction of a new growth references for china based on urban chinese children: comparison with the WHO growth standards. *PLOS ONE*. 2013 Mar 19;8(3):e59569.
- [95] Transfusion of infants and children [Internet]. [cited 2017 Sep 5]. Available from: <https://www.transfusionguidelines.org/transfusion-handbook/10-effective-transfusion-in-paediatric-practice/10-3-transfusion-of-infants-and-children>
- [96] CLINICAL PHARMACOLOGY AND BIOPHARMACEUTICS REVIEW(S). APPLICATION NUMBER: 021825Orig1s000 [Internet]. CENTER FOR DRUG EVALUATION AND RESEARCH; 2011. Available from: [https://www.accessdata.fda.gov/drugsatfda\\_docs/nda/2011/021825Orig1s000ClinPharmR.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/021825Orig1s000ClinPharmR.pdf)
- [97] Smith CT, Williamson PR, Beresford MW. Methodology of clinical trials for rare diseases. *Best Pract Res Clin Rheumatol*. 2014 Apr 1;28(2):247–62.
- [98] Sampson MR, Benjamin DK, Cohen-Wolkowicz M. Evidence-based guidelines for pediatric clinical trials: focus on StaR Child Health. *Expert Rev Clin Pharmacol*. 2012 Sep;5(5):525–31.
- [99] Russu A, van Zwet E, De Nicolao G, Della Pasqua O. Modelling of the outcome of non-inferiority trials by integration of historical data. *J Pharmacokinet Pharmacodyn*. 2011 Oct;38(5):595–612.
- [100] Mascha EJ, Sessler DI. Equivalence and noninferiority testing in regression models and repeated-measures designs. *Anesth Analg*. 2011 Mar;112(3):678–87.
- [101] Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat*. 1979;65–70.

- 
- [102] Hampson LV, Whitehead J, Eleftheriou D, Brogan P. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. *Stat Med.* 2014 Oct 30;33(24):4186–201.
- [103] Dodds MG, Hooker AC, Vicini P. Robust population pharmacokinetic experiment design. *J Pharmacokinet Pharmacodyn.* 2005 Feb 1;32(1):33–64.
- [104] Best NG, Tan KK, Gilks WR, Spiegelhalter DJ. Estimation of population pharmacokinetics using the Gibbs sampler. *J Pharmacokinet Pharmacodyn.* 1995;23(4):407–435.
- [105] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian data analysis.* Vol. 2. CRC press Boca Raton, FL; 2014.
- [106] Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput.* 2000;10(4):325–337.
- [107] WinBUGS [Internet]. MRC Biostatistics Unit. [cited 2017 Sep 12]. Available from: <https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>
- [108] Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions. *Stat Med.* 2009;28(25):3049–3067.
- [109] OpenBUGS [Internet]. [cited 2017 Sep 12]. Available from: <http://www.openbugs.net/w/FrontPage>
- [110] Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, et al. Stan: a probabilistic programming language. *J Stat Softw.* 2016;20:1–37.
- [111] Stan [Internet]. Available from: <http://mc-stan.org/>
- [112] Plummer M, others. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd international workshop on distributed statistical computing* [Internet]. Vienna, Austria; 2003. p. 125. Available from: <http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Drafts/Plummer.pdf>
- [113] JAGS - Just Another Gibbs Sampler [Internet]. Available from: <http://mcmc-jags.sourceforge.net/>
- [114] Beal S, Sheiner LB, Boeckmann A, Bauer RJ. *NONMEM User's Guides.* Icon Development Solutions; 2009.

- 
- [115] Bauer RJ, Ludden TM. Improvements and new estimation methods in NONMEM 7 for PK/PD population analysis. In 2009. Available from: [http://www.page-meeting.org/print\\_abstract.asp?abstract\\_id=1516](http://www.page-meeting.org/print_abstract.asp?abstract_id=1516)
- [116] Gilks WR, Richardson S, Spiegelhalter D. Markov chain Monte Carlo in practice [Internet]. CRC press; 1995. Available from: [https://books.google.it/books?hl=it&lr=&id=TRXrMWY\\_i2IC&oi=fnd&pg=PR16&dq=Gilks,+W.R.,+Richardson,+S.+%26+Spiegelhalter,+D.+Markov+Chain+Monte+Carlo+in+practice,+Chapman+and+Hall,+London+\(1996\).&ots=7iRptmPtsw&sig=kq\\_111\\_oFB531ap3gnsj4uNHC7k](https://books.google.it/books?hl=it&lr=&id=TRXrMWY_i2IC&oi=fnd&pg=PR16&dq=Gilks,+W.R.,+Richardson,+S.+%26+Spiegelhalter,+D.+Markov+Chain+Monte+Carlo+in+practice,+Chapman+and+Hall,+London+(1996).&ots=7iRptmPtsw&sig=kq_111_oFB531ap3gnsj4uNHC7k)
- [117] WinBUGS Development Web-site [Internet]. [cited 2017 Sep 12]. Available from: <http://winbugs-development.mrc-bsu.cam.ac.uk/>
- [118] Metrum Research Group. BUGS BUGSModelLibrary [Internet]. Available from: <https://bitbucket.org/metrumrg/bugsmode library/wiki/Home>
- [119] Metrum Research Group [Internet]. Available from: <http://metruminstitute.org/>
- [120] Metrum Research Group. Bayesian pharmacometric modeling with BUGS, NONMEM, and Stan [Internet]. Available from: <https://www.metrumrg.com/publication/bayesian-pharmacometric-modeling-bugs-nonmem-stan/>
- [121] Oberon microsystems. BlackBox [Internet]. Available from: <http://www.oberon.ch/blackbox.html>
- [122] Borella E, Carrara L, Lavezzi SM, Mezzalana E, Pasotti L, De Nicolao G, et al. Evaluation of software tools for Bayesian estimation on population models with count and continuous data. In 2015. Available from: <https://www.page-meeting.org/?abstract=3452>
- [123] Gastonguay MR, Gillespie B, Bauer RJ. Comparison of MCMC simulation results using NONMEM 7 or WinBUGS with the BUGSModelLibrary. In 2010. Available from: <https://www.page-meeting.org/?abstract=1762>
- [124] Bauer RJ. NONMEM Users Guide. Introduction to NONMEM 7.4.1. July 2017. Available from: <https://nonmem.iconplc.com/nonmem741/nm741.pdf>

- 
- [125] Bauer RJ, Guzy S, Ng C. A survey of population analysis methods and software for complex pharmacokinetic and pharmacodynamic models with examples. *AAPS J.* 2007;9(1):E60–E83.
- [126] Dokoumetzidis A, Aarons L. Propagation of population pharmacokinetic information using a Bayesian approach: comparison with meta-analysis. *J Pharmacokinet Pharmacodyn.* 2005;32(3):401–418.
- [127] Duffull SB, Kirkpatrick CM, Green B, Holford NH. Analysis of population pharmacokinetic data using NONMEM and WinBUGS. *J Biopharm Stat.* 2004;15(1):53–73.
- [128] Zhou Z, Rodman JH, Flynn PM, Robbins BL, Wilcox CK, D’Argenio DZ. Model for intracellular Lamivudine metabolism in peripheral blood mononuclear cells ex vivo and in human immunodeficiency virus type 1-infected adolescents. *Antimicrob Agents Chemother.* 2006;50(8):2686–2694.
- [129] Gueorguieva I, Aarons L, Rowland M. Diazepam pharmacokinetics from preclinical to phase I using a Bayesian population physiologically based pharmacokinetic model with informative prior distributions in WinBUGS. *J Pharmacokinet Pharmacodyn.* 2006;33(5):571–594.
- [130] Dubois VFS, de Witte WEA, Visser SAG, Danhof M, Della Pasqua O, others. Assessment of interspecies differences in drug-induced QTc interval prolongation in cynomolgus monkeys, dogs and humans. *Pharm Res.* 2016;33(1):40–51.
- [131] Lunn DJ, Best N, Thomas A, Wakefield J, Spiegelhalter D. Bayesian analysis of population PK/PD models: general concepts and software. *J Pharmacokinet Pharmacodyn.* 2002;29(3):271–307.
- [132] MRC Biostatistics Unit. PKBugs [Internet]. Available from: <http://winbugs-development.mrc-bsu.cam.ac.uk/pkbugs/home.html>
- [133] Kokash N, Moodie SL, Smith MK, Holford N. Implementing a domain-specific language for model-based drug development. *Procedia Comput Sci.* 2015;63:308–316.
- [134] DDMoRe. Interoperability Framework - Official Release [Internet]. 2016. Available from: <http://www.ddmore.eu/official-release-interoperability-framework>

- 
- [135] Harnisch L, Matthews I, Chard J, Karlsson MO. Drug and disease model resources: a consortium to create standards and tools to enhance model-based drug development. *CPT Pharmacomet Syst Pharmacol*. 2013;2(3):1–3.
- [136] DDMoRe Foundation [Internet]. Available from: <https://www.ddmore.foundation/>
- [137] Swat MJ, Moodie SL, Wimalaratne SM, Kristensen NR, Lavielle M, Mari A, et al. Pharmacometrics Markup Language (PharmML): opening new perspectives for model exchange in drug development. *CPT Pharmacomet Syst Pharmacol*. 2015;4(6):316–319.
- [138] Terranova N, Lavielle M, Smith MK, Comets E, Harling K, Nordgren R, et al. Standardized Output: flexible and tool-independent storage format of typical M&S results. In 2015. Available from: <https://www.page-meeting.org/?abstract=3599>
- [139] Smith MK, Moodie SL, Bizzotto R, Blaudez E, Borella E, Carrara L, et al. Model Description Language (MDL): A standard for modelling and simulation. *CPT Pharmacomet Syst Pharmacol* [Internet]. 2017 [cited 2017 Sep 12]; Available from: <http://onlinelibrary.wiley.com/doi/10.1002/psp4.12222/full>
- [140] R: a language and environment for statistical computing. R Foundation for Statistical Computing. [Internet]. Vienna, Austria: R Development Core Team; 2008. Available from: <http://www.R-project.org>.
- [141] Waltemath D, Wolkenhauer O. How modeling standards, software, and initiatives support reproducibility in systems biology and systems medicine. *IEEE Trans Biomed Eng*. 2016;63(10):1999–2006.
- [142] Smith L, Swat MJ, Barhak J. Sharing formats for disease models. In: *Proceedings of the Summer Computer Simulation Conference* [Internet]. Society for Computer Simulation International; 2016 [cited 2017 Sep 12]. p. 5. Available from: <http://dl.acm.org/citation.cfm?id=3015579>
- [143] DDMoRe [Internet]. Available from: <http://repository.ddmore.eu/>
- [144] Toffolo G, De Grandi F, Cobelli C. Estimation of  $\beta$ -cell sensitivity from intravenous glucose tolerance test C-peptide data: knowledge of the kinetics avoids errors in modeling the secretion. *Diabetes*. 1995;44(7):845–854.

- 
- [145] Cretti A, Lehtovirta M, Bonora E, Brunato B, Zenti MG, Tosi F, et al. Assessment of  $\beta$ -cell function during the oral glucose tolerance test by a minimal model of insulin secretion. *Eur J Clin Invest.* 2001;31(5):405–416.
- [146] Magni P, Sparacino G, Bellazzi R, Toffolo GM, Cobelli C. Insulin minimal model indexes and secretion: Proper handling of uncertainty by a bayesian approach. *Ann Biomed Eng.* 2004;32(7):1027–1037.
- [147] Van Cauter E, Mestrez F, Sturis J, Polonsky KS. Estimation of insulin secretion rates from C-peptide levels: comparison of individual and standard kinetic parameters for C-peptide clearance. *Diabetes.* 1992;41(3):368–377.
- [148] Magni P, Bellazzi R, Sparacino G, Cobelli C. Bayesian identification of a population compartmental model of C-peptide kinetics. *Ann Biomed Eng.* 2000;28(7):812–823.
- [149] Shapiro ET, Tillil H, Rubenstein AH, Polonsky KS. Peripheral insulin parallels changes in insulin secretion more closely than C-peptide after bolus intravenous glucose administration. *J Clin Endocrinol Metab.* 1988;67(5):1094–1099.
- [150] Swat MJ, Grenon P, Wimalaratne S. ProbOnto: ontology and knowledge base of probability distributions. *Bioinformatics.* 2016;32(17):2719–2721.