



Universiteit
Leiden
The Netherlands

**From data to models : reducing uncertainty in benefit risk assessment :
application to chronic iron overload in children**

Bellanti, F.

Citation

Bellanti, F. (2015, September 24). *From data to models : reducing uncertainty in benefit risk assessment : application to chronic iron overload in children*. Retrieved from <https://hdl.handle.net/1887/35437>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/35437>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/35437> holds various files of this Leiden University dissertation

Author: Bellanti, Francesco

Title: From data to models : reducing uncertainty in benefit-risk assessment : application to chronic iron overload in children

Issue Date: 2015-09-24

FROM DATA TO MODELS: REDUCING UNCERTAINTY IN BENEFIT-RISK ASSESSMENT
APPLICATION TO CHRONIC IRON OVERLOAD IN CHILDREN

Francesco Bellanti

Ph.D. Thesis, Leiden University, September 2015

FROM DATA TO MODELS: REDUCING UNCERTAINTY IN BENEFIT-RISK ASSESSMENT

Application to chronic iron overload in children

Proefschrift

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College van Promoties
te verdedigen op donderdag 24 september 2015
klokke 11:15 uur

door

Francesco Bellanti

geboren te Messina, Italië
in 1986

Promotor Prof. Dr. M. Danhof

Co-promotor Prof. Dr. O.E. Della Pasqua

Promotiecommissie Prof. Dr. P.H. van der Graaf

Prof. Dr. C.A.J. Knibbe

Prof. Dr. G. Pons, Université Paris Descartes, Parijs

Prof. Dr. J. van den Anker, Universitäts Kinderspital, Basel

Prof. Dr. J.J. Meulman, Leiden University

“If you want to go fast, go alone

If you want to go far, go together”

African proverb

a Simona

The research leading to this thesis is part of the DEferiprone Evaluation in Paediatrics (DEEP) project, supported by the FP7 Framework Research Program “HEALTH-2010.4.2-1: Off-patent medicines for children”. The research has been performed at the Division of Pharmacology of the Leiden Academic Centre for Drug Research (LACDR), Leiden University, Leiden, the Netherlands.

The printing of this thesis was financially supported by:

Quantitative Solutions B.V.

GlaxoSmithKline

Thalassaemia International Federation

Cover and layout by Daniele Bertucelli

Printed by GVO printers & designers B.V., Ede, the Netherlands

ISBN: 978-90-6464-901-1

©2015 Francesco Bellanti (francesco_bellanti@libero.it)

No part of this thesis may be reproduced or transmitted in any form or by any means without written permission of the author.

TABLE OF CONTENTS

Section I	General introduction	9
Chapter 1	Integration of PKPD relationships into Benefit-Risk Analysis	11
Chapter 2	Model-informed benefit-risk assessment of iron chelation in transfusion-dependent haemoglobinopathies - Scope and intent of the investigation	49
Section II	Optimising evidence generation in paediatric trials	65
Chapter 3	Population pharmacokinetics of deferiprone in healthy subjects	67
Chapter 4	Sampling optimisation in pharmacokinetic bridging studies: use of deferiprone in children with β -thalassaemia	87
Chapter 5	Model-based dosing recommendations for the use of deferiprone in children affected by transfusional iron overload younger 6 years of age	109
Section III	Integrated evaluation of efficacy and safety by modelling and simulation	129
Chapter 6	A disease model for iron overload in patients affected by transfusion-dependent diseases	131
Chapter 7	Model-based optimisation of deferoxamine chelation therapy	153
Chapter 8	Model-based characterisation of the acute and long-term unfavourable effects of iron chelation therapy	177
Section IV	Clinical Trial Simulations: accounting for exposure, disease progression and uncertainty in benefit-risk analysis	201
Chapter 9	Model-based evaluation of benefit-risk balance in children	203
Section V	Conclusion and perspectives	227
Chapter 10	Model-informed assessment of the benefit-risk profile of medicines for children: summary, conclusions and perspectives	229
Chapter 11	Nederlandse samenvatting (Synopsis in Dutch)	247

SECTION I

General introduction

CHAPTER 1

Integration of PKPD relationships into Benefit-Risk Analysis

F Bellanti, RC van Wijk, M Danhof, and O Della Pasqua

Br J Clin Pharmacol – 2015

Summary

Aim: Despite the continuous endeavour to achieve high standards in medical care through effectiveness measures, a quantitative framework for the assessment of the benefit-risk balance (BRB) is lacking prior to drug approval. The aim of this short review is to summarise the approaches currently available for benefit-risk assessment. In addition, we propose the use of pharmacokinetic-pharmacodynamic (PKPD) modelling as the pharmacological basis for evidence synthesis and evaluation of novel therapeutic agents.

Methods: A comprehensive literature search has been performed using MESH terms in Pubmed, in which articles describing benefit-risk assessment and modelling and simulation (M&S) were identified. In parallel, a critical review of multi-criteria decision analysis (MCDA) is presented as a tool for characterising a drug's safety and efficacy profile.

Results: A definition of benefits and risks has been proposed by the European Medicines Agency (EMA), in which qualitative and quantitative elements are included. However, in spite of the value of MCDA as a quantitative method, decisions about BRB continue to rely on subjective expert opinion. By contrast, a model-informed approach offers the opportunity for a more comprehensive evaluation of BRB before extensive evidence is generated in clinical practice.

Conclusions: BRB should be an integral part of risk management and as such considered prior to drug approval. M&S can be incorporated into MCDA to support the evidence synthesis as well evidence generation taking into account the underlying correlations between favourable and unfavourable effects. In addition, it represents a valuable tool for the optimisation of protocol design in effectiveness trials.

1.1 Benefit-Risk Analysis: the current situation

Despite the recognised implications of unmet medical needs and challenges in dealing with new diseases, the current regulatory framework in the European Union has made drug approval a demanding task. This situation is compounded by emerging safety findings, which have led to post-approval withdrawals of more than a dozen products with high therapeutic potential in the past decade (1,2). Such a landscape places regulators, clinical scientists and drug developers with yet another dilemma: how to balance rapid access to new drugs *versus* gathering comprehensive data on efficacy and safety? (3). Currently, regulators make these decisions in an isolated, fragmented, and to a large extent subjective manner.

The decision to approve a new medicinal product is based on the assumption that a systematic review of all available data provides an accurate, unbiased picture of a drug's efficacy and safety. This assumption may, however, not be true for the large majority of drugs; the evidence generated to support regulatory submission does not always account for the overall heterogeneity of the target population, the impact of treatment on disease progression or external confounding factors on treatment response. Moreover, one needs to acknowledge that the information gathered in the context of pivotal clinical trials may not provide evidence that dose selection, dosing regimen, and treatment duration are truly optimal.

Undoubtedly, efficient gathering and use of data are required to answer the clinical questions that arise with new drugs or therapeutic interventions. Among other things one needs to distinguish effectiveness from clinical response. In addition, it is crucial to understand whether there is added value, as compared with other treatments. These are multidimensional questions which require clear understanding of how data will be generated and how benefit and risk will be quantified. Whereas different theoretical considerations and techniques have been used by health technology assessment agencies, a clear framework for benefit risk (BR) assessment is still lacking during drug development and subsequently for regulatory approval. Consequently, decision making at important milestones in R&D and at submission remain empirical, inconsistent and more often than not, non-transparent(1,4–8).

In the past years awareness about the aforementioned issues has increased significantly. Several projects (9–13) have been funded to evaluate some of the available methodologies and better understand the requirements for a more systematic approach to BR analysis. In this context, the work of the Committee for Medicinal Products for Human Use (CHMP) of the European Medicines Agency (EMA) is particularly relevant. Starting in 2006, a working group was installed to examine the issue and provide recommendations about ways to improve BR assessment, including aspects such as transparency, consistency and

communication between stakeholders (9). Among the techniques evaluated by the working group, quality-adjusted life years (QALYs) and number needed to treat (NNT) were found to be the most used concepts in clinical practice, very likely due to their simplicity (9,14). However, these methods are qualitative in nature and as such lack some important features that allow one to make appropriate inferences about quantitative differences, especially when comparing treatment options. There is a clear need for more comprehensive methodologies, which enable better integration of data and facilitate the evaluation of complex clinical scenarios that arise in real life.

Most of these complexities seem to have been addressed by the development of multi-criteria decision analysis (MCDA), an integrative approach that has gained interest from the scientific and clinical community over the last few years. From 2009 to 2011, data can be found for nine products which have been evaluated by MCDA alone, or in combination with simulation, decision trees or Markov modelling (15).

In this review, a brief overview of different techniques for the evaluation of benefit and risk is presented, with especial focus on the contribution of quantitative methodologies to the development and approval of novel medicines. Two main topics are discussed initially. First, the definition of benefit-risk balance and the impact of qualitative and quantitative methodologies on the measurement of benefit and risk during the drug development process (7). In addition, we consider further refinement of the approaches used for assessing BRB by integrating it with pharmacokinetic-pharmacodynamic (PKPD) modelling. It is envisaged that modelling and simulation may account for correlations between therapeutic response and adverse events, providing a biologically plausible basis for BRB. The availability of such an integrated approach may enable better choices regarding treatment selection and dose rationale in special groups or conditions involving small numbers of patients such as rare diseases.

1.2 Methods

Initially, an exploratory literature search was performed to retrieve relevant publications to identify current quantitative approaches for benefit risk assessment (BRA) to improve decision making in drug development. Seven documents (1,4,14,22,23,61,84) were available before the exploratory phase and were used to identify 58 articles, books and reports. Based on this pool of 65 documents 21 quantitative methodologies were identified (see Table S1). This result was used to integrate the available information with a systematic literature search within PubMed, in which the name of the methodology was combined with the term benefit risk assessment/analysis, which was replaced by benefit risk or risk assessment when the query lead to an outcome of 0 publications. This resulted in 253 publications, of which 231 were rejected based on title and abstract information. The resulting 22 publications,

together with the 65 publications from the exploratory search were reviewed. Additionally, 23 publications were added on external advice, for a total number of papers reviewed in the context of current approaches for benefit risk assessment equal to 110. The steps described in this paragraph are summarised in Figure 1.

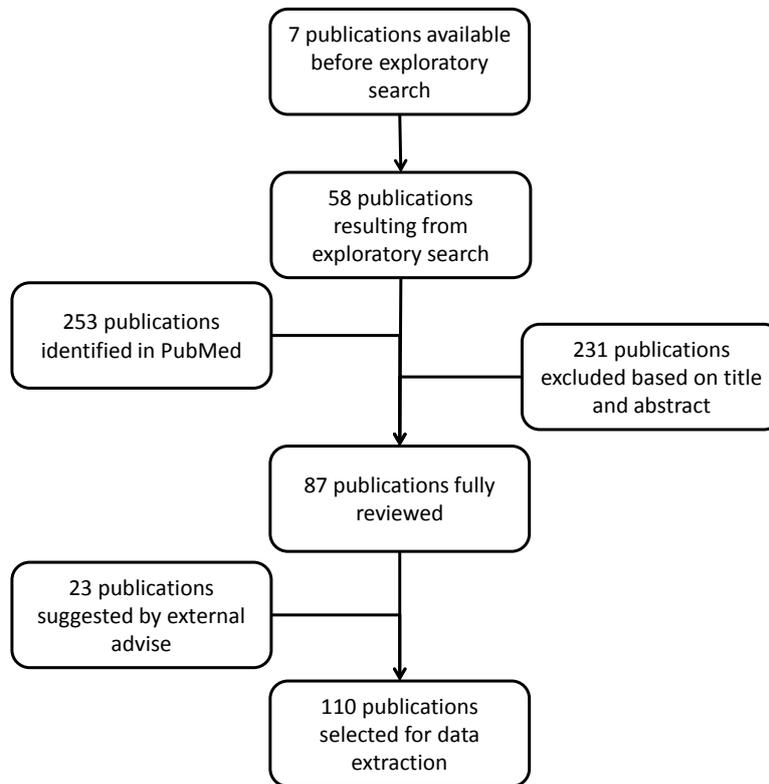


Figure 1. Flow diagram of the literature search.

1.3 Definition of benefit and risk

An important aspect of any BR analysis is the definition of both terms, and more importantly, how to measure or quantify them. Benefit is usually described as a potential effect that moves the condition of the patient from disease towards health, within a given (pre-defined) context (Table 1) (16–19). Risk is the opposite, a potential effect that moves the condition of the patient from health towards disease, also within a pre-defined context. To measure both possibilities, at least two concepts play an important role: the magnitude or severity of the effect, and its incidence or frequency. Benefit or risk is then estimated by the product of these concepts, possibly multiplied by the duration (17) or the reliability of the data (19). The BR assessment, in which the no-treatment option should not be overseen, is simply a ratio of the two components, for which pre-defined acceptance thresholds are stated.

Table 1. Glossary of terms.

Term	Definition
ADE	Adverse Drug Effects
Bayesian statistics	Probability-based statistics, concerning parameter values derived from distributions
Benefit	Favourable effect, accounting for uncertainty of that effect [as defined by the EMA]
BILAG-index	British Isles Lupus Assessment Group, a measure for severity of SLE
BR	Benefit Risk
BRAT	Benefit Risk Action Team, operating under PhRMA
CHMP	Committee for Medicinal Products in Human Use, operating under the EMA
Decision tree	Method to aid decision making by visualizing different scenarios as a series of events, and by calculating outcome based on assigned probabilities of the events
DSD	Death or serious disabled, measure of estimated outcome in the swine flu case study
EMA	European Medicines Agency
FDA	Food and Drug Administration [USA]
H1N1	Influenza virus categorized by surface proteins hemagglutinin and neuraminidase (in this case swine flu)
<i>In silico</i>	Experiment in a computer, virtually
<i>In vitro</i>	Experiments in cell cultures
<i>In vivo</i>	Experiment in animals (preclinical)
IPRED	Individual prediction, possible outcome of PKPD modelling prediction variables and parameter values of an individual patient
M&S	Modelling and Simulation, in pharmacology a way of describing data by constructing a validate model and simulate new data, as a virtual experiment
Markov model	Quantitative method of modelling states and transitions between states
MCDA	Multi criteria decision analysis, quantitative method analysing single weighted components of a problem before reassembling it to aid a final decision
NDA	New drug application, to be submitted to the FDA for approval before market access

NNH	Number needed to harm, measure of the number of patients that has to be treated to present a single adverse effect
NNT	Number needed to treat, measure of the number of patients that has to be treated to prevent a single occurrence
PhRMA	Pharmaceutical Research and Manufacturers of America
PKPD	Pharmacokinetics and pharmacodynamics, two disciplines within pharmacology concerning what the body does to the drug and what the drug does to the body, respectively
PrOACT-URL	Qualitative framework by Hammond, Keeney and Raiffa, consisting of Problem, Objective, Alternatives, Consequences, Trade-offs, Uncertainty, Risk tolerance and Linked decisions
QALY	Quality-adjusted life year, measuring the outcome of therapy by the adjustment of a quality life year, in which the patient can fully function (economically)
Risk	Unfavourable effect, accounting for uncertainty of that effect [as defined by the EMA]
RV-NNT	Relative Value adjusted number needed to treat, a type of NNT accounting for patient preference as value function
SLE	Systemic Lupus Erythematosus, an autoimmune disease
SLEDAI	DLE Disease Activity Index, a measure of severity of SLE
TURBO	Transparent Uniform Risk-Benefit Overview

Currently a slightly different definition of benefit and risk has been adopted by the EMA. They are defined respectively, as favourable and unfavourable effects and are at the same time coupled to the uncertainty of both effects (Figure 2) (14). Whereas the reasoning seems intuitive, this situation represents a mathematical challenge, i.e., integrating terms or factors that are measured in incommensurable units and in different time scales. Any reliable product of these factors imposes data manipulation or transformation to ensure that all terms are expressed in the same unit and time scale. However, the illusion of this mathematical precision tends to hide another important conceptual challenge: what is acceptable? (18). This depends on the perception and values of the stakeholders, i.e., the regulator, the clinical experts, and the patients. Procedures have been devised to ensure that perceived benefits and risks are quantified in a systematic manner. This process is known as prior elicitation and involves expert judgment. It is aimed at making subjective opinions more consistent, comprehensive and transparent (16,19,20).

Favourable effects	Uncertainty of favourable effects
Unfavourable effects	Uncertainty of unfavourable effects

Figure 2. EMA's definition of benefit and risk, where favourable effects are beneficial to the population and unfavourable effects are undesirable for the population. Uncertainty is caused by variation, biased data, limitations of data or methodology etc. Based on [14].

1.4 Current approaches

The assessment of benefit and risk has evolved in a rather empirical manner and still relies on subjective criteria, in that perceived benefits and risks depend on the context in which the treatment is used, i.e., which standard of care is set as reference and whether short and long term consequences of the intervention are considered against the progression of disease and any correlated co-morbidities or complications. Irrespective of the lack of consensus on how to assess and weight any measures associated with benefit and risk, one needs to consider two different dimensions of the problem. First, a qualitative approach is required to allow for explicit contextualisation of the problem. It is crucial to fully understand the main issues before any quantitative analysis starts, i.e., to identify the factors that contribute and/or determine benefit and risk as well as capture the views and differences of opinion from different stakeholders, especially with regard to the perception of risk, in terms of its incidence, severity, chronicity and reversibility. Second, a quantitative approach is needed in which results from the initial (qualitative) evaluation are normalised by means of mathematical and statistical procedures. Such a normalisation implies the availability of sufficient data for those endpoints and measures which have higher weights. It also imposes clear understanding of the trade-offs between benefit and risk, especially of the correlations between outcomes. Whereas these requirements seem obvious, little attention has been paid to the biological or pharmacological basis that determines treatment outcome, i.e. how exposure-response (PKPD) relationships underpin favourable and unfavourable events.

The next paragraphs will provide an overview of the available techniques, including recent examples in which benefit and risk have been evaluated in the context of regulatory

approval and treatment optimisation. Additional details of the methodologies can be found in the supplementary material (Supplementary Figure 1 and 2 and supplementary Table 1 and 2, see Appendix).

Qualitative approaches

A qualitative framework is essential for characterising benefit and risk, as it structures the problem and its context, before any actual assessments are made. It provides clarity about the possible outcomes of the assessment, as well as the input and the process in between, for example by defining which decision criteria are to be used. This framework ensures that no alternative measures or trade-offs are overlooked during the subsequent steps, i.e., during which quantitative methods are applied.

PhRMA BRAT

The Pharmaceutical Research and Manufacturers of America (PhRMA) assigned a Benefit Risk Action Team (BRAT) to create a decision framework. Their framework consists of six steps which are developed and implemented prior to drug approval. Before phase III, focus is given to the definition of a decision frame, identification of relevant outcomes, identification of the data sources, and customisation of the framework for BR analysis. At the time of filing and NDA review, attention is paid to the outcome itself as well as to the quantification and interpretation of key BR metrics (13,14). It should be noted that this framework seems to end with the decision and defence after which a drug is approved or rejected. It does not involve post marketing data, which are known to potentially change BR balance.

EMA PrOACT

The qualitative framework suggested by the EMA is based on Hammond's, Keeney's and Raiffa's PrOACT approach (21), combined with the less known addition of the so-called URL: Problem, Objective, Alternatives, Consequences, Trade-offs, Uncertainty, Risk tolerance and Linked decisions. In this way, the problem is clearly structured and information can be gathered in a consistent way to assist the decision-making process (14). Despite its general nature, the use of PrOACT-URL has proven its success since 1999. In contrast to PhRMA BRAT, the inclusion of uncertainty paves the way for a more statistically sound implementation.

Quantitative approaches

The use of a qualitative framework for assessing benefit and risk may be sufficient when complexity is minimal. This is however not the case in drug development where very complex scenarios arise. To include all data and present a sound overview of all alternatives, consequences and trade-offs, as well as differentiate between objectives otherwise

considered comparable, one or more quantitative techniques are required (1,4,11,14,17,19,22,23). A qualitative framework will still be essential to define the problem and the objectives of the analysis and as such will precede the implementation of a quantitative BR analysis.

In the past decades several methodologies have been developed and used to evaluate the BR balance of a number of drugs. These methodologies present completely different features and their use has been tailored for very specific cases, contributing to an increase in the number of options available when starting an analysis. These specificities have however made them unsuitable for subsequent application in a general BR framework. An overview of these methods (1,4,17,19,22–102, 120-124), including advantages and limitations is provided in Supplementary table 1. By contrast, multi criteria decision analysis (MCDA) in combination with decision trees has been suggested as a plausible quantitative approach that embeds the needed features for a generalised and structured framework for BR evaluation.

MCDA presents several advantages compared to other methodologies: the main one is the simplification of a complex problem by breaking it into smaller pieces and making them comparable by weighting their scores on a single scale; normalizing the different criteria allows comparison on the same ground. In addition, the uncertainty carried by the subjective component, is further reduced by the possibility of performing sensitivity analysis, in which the model provides different outcomes depending on weights variation. There are, however, still limitations. Given the complexity of the scenarios analyzed, it is often expected to observe correlations between the endpoints considered. This is not yet taken into account within the methodology, where each endpoint is analyzed in an independent manner. In the SLE-case, which is discussed in the supplementary material, the immunosuppressive effect of Benlysta and the incidence of infection might very well be correlated in a nonlinear way. This might influence the outcome, leading to biased results. Furthermore, it is a matter of concern how the input data for the decision model is provided. This is not a direct limitation of the methodology, but of how the analysis is implemented. Many quantitative methods are limited by statistics and inclusion of uncertainty, confounding factors, or limited data. The latter concerns both the experimental data, as well as preference values of different stakeholders required for weighting criteria (1). MCDA offers a statistical sound method, where probability and uncertainty are combined with preference. Its limitation lies in the complexity of data required, which is often unavailable, as well as in the subjective judgement that is required and the dependence on risk perception differences. Besides, sequential decisions require data gathering over a longer time period, especially in conditional approval (103).

Despite the aforementioned advantages, MCDA, like any other quantitative method, still relies on subjectivity. This is partly overcome by structuring the analysis in a transparent, consistent manner and by incorporating communication with different stakeholders as a critical step (14,15). In fact, communication with different stakeholders is also accounted for in NNT/NNH. Although applicability of the former to BR assessment in general is very limited because of the lack of preference data, as well as the limited statistical power (57,58), it shows an important issue in communication. Individual patients seem unable to objectively estimate their own chances. In a distribution of 1 out of 20, all 20 patients expect to be the exception, when it comes to a beneficial effect, but not in case of an adverse effect. As a result, the magnitude of risk is misperceived, as the chances of common consequences are underestimated and those of rare consequences are overestimated (8). This problem of risk perception is essential when considering including different stakeholders. Although MCDA does present data in a transparent and consistent way, it is not a technical process, but an effective design of the social processes required for subjective weighting (41).

1.5 Integration of PKPD modelling into BR analysis

Modelling and Simulation (M&S) techniques represent an invaluable resource for drug development. Of relevance for BR analysis is the opportunity that PKPD modelling offers in terms of describing variability in a parametric manner. This allows the characterization and prediction of the time course of treatment response at individual level under physiological and pathological conditions (104,105). The current emphasis on mechanism-based modelling has also the advantages of increased understanding about drug-specific and system-specific properties such as, target site distribution, binding, pharmacokinetic interactions, transduction of signals, pharmacodynamic interactions, homeostatic feedback, tolerance and disease progression (106–108). In addition, model-based simulations can provide insight into conditions that may not have been tested experimentally, unravelling patterns or responses that may represent clinically relevant changes in the BR balance.

From a technical, scientific point of view, M&S ensures for integration of data and knowledge in a continuous, objective and reproducible manner, thereby enhancing the quality of decision making (105). Over the last decade, regulatory perception and role of M&S in drug development has changed. Its relevance in clinical development has been acknowledged and processes are in place to support a more structured use of M&S (106,109).

In the next paragraphs we evaluate how the integration of M&S can be advantageous to further improve the existing framework for the evaluation of benefit-risk balance, as suggested by the EMA. To this purpose, we consider three main aspects, namely, the optimisation of evidence that is generated by clinical trials, evaluation of virtual scenarios

and mechanism-based multivariate analysis. The optimisation of the input data available for decision making entails not only the integration of data from different trials, but also the use of optimality concepts for the design of prospective clinical studies. The availability of an integrated model allows for the creation of virtual experiments, which provide a more coherent, biologically plausible basis for performing interpolations and extrapolations. In contrast to current practice, multivariate modelling allows one to establish correlations between therapeutic and adverse events of interest, which are often linked by the very pharmacological nature of the treatment. This overview is complemented by a brief discussion of the issues associated with prior elicitation, which could be better guided by the use of models, rather than empirical distributions. As such, a model-based approach could provide somewhat less subjective weighting and preferences.

Optimizing input data: M&S techniques can be used to optimize the input data available for the BR analysis. PKPD modelling allows the creation of a framework that can be refined and improved throughout the development process, by integrating data from different sources as well as by pooling the information gathered across different phases of development. This iterative process allows one to understand and distinguish drug from system-specific properties. Most importantly, it allows one to identify sources of variation and assess the clinical implications thereof. Among other things, BR analysis could be performed with and without the residual variability or in by inclusion of variability in a stepwise manner. In other words, these procedures increase the value of data whilst decreasing uncertainty (106). On the other hand, M&S can also be used to optimise the design of prospective clinical trials. The quality of the information collected can be considerably improved through optimal design (110–112), enabling the generation of more informative data input for the decision analysis. This is particularly important in special populations where limited evidence is generated, such as in paediatric diseases (106,113,114). The assumptions about the informative value of data obtained from randomised clinical trials are often overlooked. It is assumed that the output or results from a trial are consequence of the drug treatment, rather than the consequence of the interaction between drug properties, disease processes, patient characteristics and experimental protocol.

Evidence from virtual scenarios: A second aspect that could be beneficial for the BR assessment is the use of PKPD modelling for simulation purposes. The availability of a qualified or validated model may provide the opportunity to perform virtual experiments. This allows one to explore scenarios that have not been evaluated during clinical development. Not only efficacy and safety data can be considered, but also the influence of covariates such as disease severity, co-medications, co-morbidities and drug compliance can be evaluated. By inter- or extrapolating, new input data can be generated for a different

population or different dosing regimens. As such these simulated results can be subsequently used as input for BR analysis. As mentioned previously, PKPD modelling may have an even larger impact when considering special populations (114–117).

Correlating multiple endpoints: Thus far we have highlighted the fact that PKPD modelling may reduce the uncertainty in a BR analysis by optimising the information used as input. M&S techniques may overcome another important limitation of BR methodologies, namely the assumption that favourable and unfavourable events are clinically, pharmacologically and statistically independent from each other. This assumption violates our current understanding of the nature and cause of adverse events. Hence, any analysis involving multiple endpoints in a multidimensional system will have to account for the correlations between them. Moreover, we believe that these correlations are often non-linear, requiring some advanced statistical techniques to ensure that interactions between variables and covariate factors are captured accordingly. Multidimensional models can be used to assess quantitatively how endpoints are linked together and how response changes with changes in drug exposure (24).

Advantages from the integration of M&S techniques to BR analysis are not only conceptual. From a technical perspective, PKPD models may contribute to bias reduction during prior elicitation. In addition, it may provide a stronger basis for sensitivity analysis. Although weighting is a subjective procedure, expert opinions can be modelled using prior elicitation. Moreover, if the uncertainty associated with the weights is assessed, it is possible to factor in the impact of each expert's opinion on the overall analysis. Other possibilities exist to weight the experts input, by scaling their precision based on training and experience, or by assigning them to groups of thought that are more or less representative of the common opinion (26,63). PKPD models describing the underlying disease processes as well as the impact of treatment over time through virtual scenarios may facilitate prior elicitation, providing systematic, consistent input for the evaluation of weights and uncertainties. An example of the impact of M&S concepts on BR analysis is given in Table 2.

Table 2. Impact of M&S on the MCDA approach is visualised and further elucidated by the example of Benlysta. It shows clearly the emphasis on the first part of the methodology; the input data and earlier data evaluation for correlations between parameters and outcomes.

MCDA	Modelling & Simulation	Example: Benlysta
Step 0: Input data gathering	Step 0.1: Explore and refine the informative contents of data, accounting for variability and uncertainty. Step 0.2: Incorporation of virtual measurements (samples), by evidence generation through simulations.	Step 0.1: Distinguish between-subject variability in relevant parameters from residual error. Step 0.2: Evaluate parameter uncertainty by exploring the implications of different experimental protocol conditions.
Step 1: Defining decision context	Step 1.1: Prioritising elements which affect variability and /or uncertainty.	
Step 2: Identifying options	Step 2.1: Inference by extrapolation, e.g., an additional arm that has not been tested clinically.	Step 2.1: Assess treatment response for alternative dosing regimens than the actual treatment arms in the trial (i.e., 1 and 10 mg doses)
Step 3: Identifying objectives and criteria	Step 3.1: Assess outcomes taking into account the correlation between events.	Step 3.1: The correlation between immunosuppressive effects and incidence of infection can be incorporated into the model, enabling accurate evaluation of the impact of different dose levels on outcome.
Step 4: Scoring	Step 4.1: Estimation of the correlation between events in a parametric manner, thereby avoiding biased scoring of the data.	Step 4.1: Estimation of the parameters describing the nonlinear relationship between immunosuppressive effects and incidence of infection in patients undergoing long term treatment.
Step 5: Weighting factors for differences of opinion	Step 5.1: Prior elicitation of expert opinions can be translated into consistent weighting, including distributions describing differences of opinion (e.g., priors in parameter distributions).	Step 5.1: Simulate outcomes for Benlysta-treated patients taking into account different weighting factors.
Step 6: Combining data	Step 6.1: Simulated scenarios increase the quality of the data and therefore the quality of the overall value, by increasing	Step 6.1: Simulation of different treatment arms to explore the implications of dose selection.

	granularity.	
Step 7: Examining data	Step 6.1: Outcome evaluation is not limited to the data, but to evidence arising from virtual clinical trials, including patients who belong to risk groups (e.g., those who meet exclusion criteria)	
Step 8: Sensitivity analysis	Step 8.1: Irrespective of the decision criteria, model parameters on which the data are based can also be analysed.	Step 8.1: The PKPD model of Benlysta has been evaluated by sensitivity analysis.

1.6 Discussion and conclusion

In this short review, an overview was given of the methodologies currently used for the evaluation of BR balance. Growing consensus suggests that a combined approach involving qualitative and quantitative methods is required to ensure meaningful evaluation and interpretation of benefit and risk data. In fact, this is recommended by the EMA, which suggests the use of ProACT-URL and MCDA.

Even though a more structured approach is still lacking for BR analysis, MCDA seems to address the need for a multidimensional characterisation of the scenarios that arise in drug development and in the clinical practice. One of its limitations is the way uncertainty is handled; there is the need to further reduce the uncertainty or preferably to capture it accordingly. Attempts have been made to construct stochastic multi-attribute models, also known as stochastic multi-criteria acceptability analysis (SMAA), which incorporates uncertainty regarding the criteria measurements. SMAA provides the possibility to include the sampling variation and to characterize typical trade-offs supporting a drug BR profile without knowing or eliciting the (exact numerical) preferences beforehand (119). An analysis without preference information is valuable when preferences cannot be elicited or when the potential benefits of a drug have to be assessed across a wide range of preferences. This latter situation occurs, for example, when different subgroups of patients are considered. However, stochastic methods do not eliminate discrepancies between perceived risk or benefit and their biological and pharmacological plausibility. Undoubtedly, integration of mechanism-based modelling to multi-criteria decision methods will enhance our ability to characterise benefit-risk balance. It will provide indirect evidence from virtual scenarios in a more effective manner than sensitivity analysis and other statistical techniques have allowed for. Such an integrated approach will also represent an advancement for the field of modelling and simulation, which is often restricted to single endpoints, facilitating the assessment of causality and correlation between favourable and unfavourable events (118). Unfortunately, in literature there are very few examples that present in a clear manner the concepts discussed throughout this manuscript. Among them though, two publications

provide an excellent illustration of these concepts: the work carried out by Bender et al (125) shows how exposure-response relationships quantified through model-based approach for multiple endpoints can be used to explore and assess BR across different dosing regimens in the context of oncology trials. In the same way, the work by Pink et al (126) shows the feasibility of integrating M&S with pharmacoeconomic analysis to inform decision making throughout the whole drug development process and possibly achieve personalised evaluations. Both examples support the fact that PKPD relationships are crucial in the assessment of a drug efficacy and safety and should not be omitted when performing a BR appraisal.

In addition, we propose here the use of PKPD modelling as the pharmacological basis for evidence synthesis and evaluation of novel therapeutic agents. Various methodologies are available for evidence synthesis, and among them network meta-analysis (NMA) has been widely used in BR analyses to combine all available evidence (127-128). These approaches though, rely on very large amount of information and as discussed in this manuscript depend only on the evidence generated. As opposite to a model-based approach, they are not able to provide an understanding or a quantification of the underlying PKPD mechanisms and subsequently cannot be used to anticipate and explore virtual scenarios through Clinical Trial Simulations and/or Not-in-trial Simulations (129). In a post-marketing phase the contribute of NMA is indeed invaluable but in a pre-marketing evaluation where limited data is available PKPD cannot be ignored and to our understanding may be crucial for a comprehensive BR evaluation.

In conclusion, it should be highlighted that models do not make decisions, people do. Ultimately, patients, clinicians, drug developers and regulators need to acknowledge that decisions are better made when data are presented and communicated in a clear, systematic manner. PKPD modelling can complement evidence generation by providing stakeholders the opportunity to explore conditions that have not been experimentally tested at the time of BR analysis. Regardless of the limitations models and simulation scenarios may have, model-based evaluation is likely to outperform gut feeling, which often prevails in clinical decision-making.

References

1. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. *Value Health*. 2010 Aug;13(5):657–66.
2. Busch M, Walderhaug M, Custer B, Allain J-P, Reddy R, McDonough B. Risk assessment and cost-effectiveness/utility analysis. *Biol J Int Assoc Biol Stand*. 2009;37:78–87.
3. Eichler H-G, Pignatti F, Flamion B, Leufkens H, Breckenridge A. Balancing early market access to new drugs with the need for benefit/risk data: a mounting dilemma. *Nat Rev Drug Discov*. 2008;7:818–26.
4. Holden WL. Benefit-risk analysis : a brief review and proposed quantitative approaches. *Drug Saf*. 2003 Jan;26(12):853–62.
5. Bjornson D. Interpretation of drug risk and benefit: individual and population perspectives. *Ann Pharmacother*. 2004;38:694–9.
6. Okie S. Safety in Numbers — Monitoring Risk in Approved Drugs. *N Engl J Med*. 2005;352:1173–6.
7. Califf R. Benefit assessment of therapeutic products: the Centers for Education and Research on Therapeutics. *Pharmacoepidemiol Drug Saf*. 2007;16(5-16).
8. Cohen J, Neumann P. What's more dangerous, your aspirin or your car? Thinking rationally about drug risks (and benefits). *Heal Aff*. 2007;26:636–46.
9. EMA. Benefit-risk methodology project - Project description. 2009.
10. Liberti B, McAuslane N, Walker S. Progress on the Development of a Benefit / Risk Framework for Evaluating Medicines. *Regul Focus*. 2010;15:32–7.
11. FDA. Structured Approach to Benefit-Risk Assessment in Drug Regulatory Decision-Making - Draft PDUFA V Implementation Plan. 2013.
12. Walker S, McAuslane N, Liberti L, Salek S. Measuring benefit and balancing risk: strategies for the benefit-risk assessment of new medicines in a risk-averse environment. *Clin Pharmacol Ther*. 2009;85:241–6.
13. Coplan P, Noel R, Levitan B, Ferguson J, Mussen F. Development of a Framework for Enhancing the Transparency , Reproducibility and Communication of the Benefit – Risk Balance of Medicines. *Clin Pharmacol Ther*. 2011;89:312–5.
14. EMA. Benefit-risk methodology project - work package 2 report: Applicability of current tools and processes for regulatory benefit-risk assessment. 2011 p. 1–33.

15. Phillips L. The European Medicines Agency Benefit-Risk Project (presentation). 2013.
16. Lane D, Hutchinson T. The Notion of “Acceptable Risk”: The Role of Utility in Drug Management. *J Chronic Dis.* 1987;40:621–5.
17. Edwards IR, Wiholm BE MC. Concepts in Risk-Benefit Assessment. A Simple Merit Analysis of a Medicine? *Drug Saf.* 1996;15(1):1–7.
18. Ernst E, Resch K. Risk-Benefit Ratio or Risk-Benefit Nonsense? *J Clin Epidemiol.* 1996;49:1203–4.
19. Beckmann J. Basic aspects of risk-benefit analysis. *Semin Thromb Hemost.* 1999 Jan;25(1):89–95.
20. Morgan G, Henrion M. *Uncertainty - A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press; 1990.
21. Hammond J, Keeney R, Raiffa H. *Smart Choices.* Harvard Business School Press; 1999.
22. Curtin F, Schulz P. Assessing the benefit:risk ratio of a drug - randomized naturalistic evidence. *Dialogues Clin Neurosci.* 2011;13:183–90.
23. Puhan MA, Singh S, Weiss CO, Varadhan R, Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. *BMC Med Res Methodol.* *BMC Medical Research Methodology;* 2012 Jan;12(1):173.
24. Ahn JE, French JL. Longitudinal aggregate data model-based meta-analysis with NONMEM: approaches to handling within treatment arm correlation. *J Pharmacokinet Pharmacodyn.* 2010 Apr;37(2):179–201.
25. Airoldi M, Morton A. Adjusting life for quality or disability: stylistic difference or substantial dispute? *Health Econ.* 2009;18:1237–47.
26. Albert I, Donnet S, Guihenneuc-Jouyaux C, Low-Choy S, Mengersen K, Rousseau J. Combining Expert Opinions in Prior Elicitation. *Bayesian Anal.* 2012 Sep;7(3):503–32.
27. Ashby D, Smith AF. Evidence-based medicine as Bayesian decision-making. *Stat Med.* 2000 Dec 15;19(23):3291–305.
28. Barrett J, Skolnik J, Jayaraman B, Patel D, Adamson P. Discrete event simulation applied to pediatric phase I oncology designs. *Clin Pharmacol Ther.* 2008 Dec;84(6):729–33.
29. Berchiolla P, Scarinzi C, Snidero S, Gregori D. Comparing models for quantitative risk assessment: an application to the European Registry of foreign body injuries in children. *Stat Methods Med Res.* 2013 Feb 19.

CHAPTER 1

30. Brass EP, Lofstedt R, Renn O. Improving the decision-making process for nonprescription drugs: a framework for benefit-risk assessment. *Clin Pharmacol Ther.* Nature Publishing Group; 2011 Dec;90(6):791–803.
31. Janssen M, Over J, van der Poel C, Cuijpers H, van Hout B. A probabilistic model for analyzing viral risks of plasma-derived medicinal products. *Transfusion.* 2008;48:153–62.
32. Chauvin P, Josselin J-M, Heresbach D. Incremental net benefit and acceptability of alternative health policies: a case study of mass screening for colorectal cancer. *Eur J Health Econ.* 2012 Jun;13(3):237–50.
33. Chawla A, Mytelka D, McBride S, Nellesen D, Elkins B, Ball D, Kalsekar A, Towse A, Garrison LP. Estimating the incremental net health benefit of requirements for cardiovascular risk evaluation for diabetes therapies. *Pharmacoepidemiol Drug Saf.* 2014;23(3):268–77.
34. Chuang-Stein C. A new proposal for benefit-less-risk analysis in clinical trials. *Control Clin Trials.* 1994 Feb;15(1):30–43.
35. CIOMS working group I. Benefit-Risk Balance for Marketed Drugs: Evaluating Safety Signals. Report of CIOMS Working Group IV. 1998.
36. Citrome L, Ketter TA, Cucchiaro J, Loebel A. Clinical assessment of lurasidone benefit and risk in the treatment of bipolar I depression using number needed to treat, number needed to harm, and likelihood to be helped or harmed. *J Affect Disord.* Elsevier; 2014 Feb;155:20–7.
37. Corey-Lisle PK, Peck R, Mukhopadhyay P, Orsini L, Safikhani S, Bell JA, Hortobagyi G, Roche H, Conte P, Revicki DA. Q-TWiST analysis of ixabepilone in combination with capecitabine on quality of life in patients with metastatic breast cancer. *Cancer.* 2012 Jan 15;118(2):461–8.
38. Corso PS, Hammitt JK, Graham JD, Dicker RC, Goldie SJ. Assessing Preferences for Prevention Versus Treatment Using Willingness to Pay. *Med Decis Mak.* 2002 Oct 1;22(5 suppl):S92–S101.
39. DeCosse JJ, Cennerazzo WJ. A quality-adjusted time without symptoms or toxicity (Q-TWiST) analysis of adjuvant radiation therapy and chemotherapy for resectable rectal cancer. *J Natl Cancer Inst.* 1996 Nov 20;88(22):1686.
40. Djulbegovic B, Hozo II, Fields K, Sullivan D. High-Dose Chemotherapy in the Adjuvant Treatment of Breast Cancer: Benefit/Risk Analysis. *Cancer Control.* 1998;5(5):394–405.
41. Dogson. Multi-criteria analysis: a manual. Communities and Local Government Publications; 2009.
42. Donovan AK, Smith KJ, Ragni M V. Anticoagulation duration in heterozygous factor V Leiden: a decision analysis. *Thromb Res.* Elsevier Ltd; 2013 Dec;132(6):724–8.
43. Earnshaw SR, Brogan AP, McDade CL. Model-based cost-effectiveness analyses for prostate cancer chemoprevention : a review and summary of challenges. *Pharmacoeconomics.* 2013 Apr;31(4):289–304.

44. EMA. Benefit-risk methodology project Work package 4 report: Benefit-risk tools and processes. 2012 p. 1–20.
45. Eriksen S, Keller LR. A Multiattribute-utility-function Approach to Weighing the Risks and Benefits of Pharmaceutical Agents. *Med Decis Mak.* 1993 Jun 1;13(2):118–25.
46. Fahey T, Griffiths S, Peters TJ. Evidence based purchasing: understanding results of clinical trials and systematic reviews. *BMJ.* 1995 Oct 21;311(7012):1056–60.
47. Felli JC, Noel RA, Cavazzoni PA. A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives. *Med Decis Making.* 2009;29(1):104–15.
48. Francis GS. Importance of benefit-to-risk assessment for disease-modifying drugs used to treat MS. *J Neurol.* 2004 Sep;251 Suppl v42–v49.
49. Gelber RD, Gelman RS, Goldhirsch a. A quality-of-life-oriented endpoint for comparing therapies. *Biometrics.* 1989 Sep;45(3):781–95.
50. Gelber RD, Cole BF, Gelber S, Goldhirsch A. Comparing Treatments Using Quality-Adjusted Survival : The Q-TWiST Method. *Am Stat.* 2014;49(2):161–9.
51. Geva A, Gray J. A quantitative analysis of optimal treatment capacity for perinatal asphyxia. *Med Decis Making.* 2012;32(2):266–72.
52. Greving JP, Buskens E, Koffijberg H, Algra A. Cost-effectiveness of aspirin treatment in the primary prevention of cardiovascular disease events in subgroups based on age, gender, and varying cardiovascular risk. *Circulation.* 2008 Jun 3;117(22):2875–83.
53. Gupta SK. Use of Bayesian statistics in drug development: Advantages and challenges. *Int J Appl basic Med Res.* 2012 Jan;2(1):3–6.
54. Hauser JR. *Conjoint Analysis , Related Modeling , and Applications.* Advances in Marketing Research: Progress and Prospects. 2002.
55. Hirsch GB, Edelstein BL, Frosh M, Anselmo T. A simulation model for designing effective interventions in early childhood caries. *Prev Chronic Dis.* 2012 Jan;9(1):E66.
56. Hodder SC, Edwards MJ, Brickley MR, Shepherd JP. Multiattribute utility assessment of outcomes of treatment for head and neck cancer. *Br J Cancer.* 1997 Jan;75(6):898–902.
57. Holden WL, Juhaeri J, Dai W. Benefit-risk analysis: examples using quantitative methods. *Pharmacoepidemiol Drug Saf.* 2003 Dec;12(8):693–7.
58. Holden WL, Juhaeri J, Dai W. Benefit-risk analysis: a proposal using quantitative methods. *Pharmacoepidemiol Drug Saf.* 2003;12(7):611–6.

59. Hynes N, Sultan S. A prospective clinical, economic, and quality-of-life analysis comparing endovascular aneurysm repair (EVAR), open repair, and best medical treatment in high-risk patients with abdominal aortic aneurysms suitable for EVAR: the Irish patient trial. *J Endovasc Ther.* 2007 Dec;14(6):763–76.
60. Ijzerman MJ, van Til JA, Bridges JFP. A comparison of analytic hierarchy process and conjoint analysis methods in assessing treatment alternatives for stroke rehabilitation. *Patient.* 2012 Jan;5(1):45–56.
61. Inn G, Washington DC, Philips L, Cone M. Benefit-Risk Assessment Model for Medicines : Developing a Structured Approach to Decision Making Report of the Workshop organised by the CMR International Institute for Regulatory Science Stuart Walker. 2006.
62. Khan A a, Perlstein I, Krishna R. The use of clinical utility assessments in early clinical development. *AAPS J.* 2009 Mar;11(1):33–8.
63. Kinnersley N, Day S. Structured approach to the elicitation of expert beliefs for a Bayesian-designed clinical trial: a case study. *Pharm Stat.* 2013;12(2):104–13.
64. Kopylev L, Chen C, White P. Towards quantitative uncertainty assessment for cancer risks: central estimates and probability distributions of risk in dose-response modeling. *Regul Toxicol Pharmacol.* 2007 Dec;49(3):203–7.
65. Korsan B, Dykstra K, Pullman W. A clinical utility index (CUI) openly evaluates a product ' s attributes — and chance. *Pharm Exec.* 2005;
66. Krishna R. Model-based benefit-risk assessment: can Archimedes help? *Clin Pharmacol Ther.* 2009 Mar;85(3):239–40.
67. Langley RG, Strober BE, Gu Y, Rozzo SJ, Okun MM. Benefit-risk assessment of tumour necrosis factor antagonists in the treatment of psoriasis. *Br J Dermatol.* 2010 Jun;162(6):1349–58.
68. Leil TA, Feng Y, Zhang L, Paccaly A, Mohan P, Pfister M. Quantification of apixaban's therapeutic utility in prevention of venous thromboembolism: selection of phase III trial dose. *Clin Pharmacol Ther.* Nature Publishing Group; 2010 Sep;88(3):375–82.
69. Levitan B, Yuan Z, Turpie AGG, Friedman RJ, Homering M, Berlin JA, Berkowitz SD, Weinstein RB, DiBattiste PM. Benefit-risk assessment of rivaroxaban versus enoxaparin for the prevention of venous thromboembolism after total hip or knee arthroplasty. *Vasc Health Risk Manag.* 2014 Jan;10:157–67.
70. Loftus E V, Johnson SJ, Wang S-T, Wu E, Mulani PM, Chao J. Risk-benefit analysis of adalimumab versus traditional non-biologic therapies for patients with Crohn's disease. *Inflamm Bowel Dis.* 2011 Jan;17(1):127–40.
71. Lynd LD, Marra CA, Najafzadeh M, Sadatsafavi M. A quantitative evaluation of the regulatory assessment of the benefits and risks of rofecoxib relative to naproxen : an application of the incremental net-benefit framework y. *Pharmacoepidemiol Drug Saf.* 2010;19(July):1172–80.

72. Lynd LD, Najafzadeh M, Colley L, Byrne MF, Willan AR, Sculpher MJ, Johnson FR, Hauber AB. Using the incremental net benefit framework for quantitative benefit-risk analysis in regulatory decision-making--a case study of alosetron in irritable bowel syndrome. *Value Health*. 2010;13(4):411–7.
73. Lynd LD, O'Brien BJ. Advances in risk-benefit evaluation using probabilistic simulation methods: an application to the prophylaxis of deep vein thrombosis. *J Clin Epidemiol*. 2004 Aug;57(8):795–803.
74. Mann J, Ellis S, Waternaux C, Liu X, Oquendo M, Malone K. Classification Trees Distinguish Suicide Attempters in Major Psychiatric Disorders: A Model of Clinical Decision Making. *J clin psychiatry*. 2013;69(1):23–31.
75. Marino P, Roché H, Moatti J-P. High-dose chemotherapy for patients with high-risk breast cancer: a clinical and economic assessment using a quality-adjusted survival analysis. *Am J Clin Oncol*. 2008 Apr;31(2):117–24.
76. Meyboom R, Egberts A. Comparing therapeutic benefit and risk. *Therapie*. 1999;54(1):29–34.
77. Mounier N, Haioun C, Cole BF, Gisselbrecht C, Sebban C, Morel P, Marit G, Bouabdallah R, Ravoet C, Salle G, Reyes F, Lepage E. Quality of life-adjusted survival analysis of high-dose therapy with autologous bone marrow transplantation versus sequential chemotherapy for patients with aggressive lymphoma in first complete remission. *Groupe d'Etude les Lymphomes de l'Adulte (GELA). Blood*. 2000 Jun 15;95(12):3687–92.
78. Mudd PN, Groenendaal H, Bush M a, Schmith VD. Probabilistic risk analysis: improving early drug development decision making. *Clin Pharmacol Ther*. Nature Publishing Group; 2010 Dec;88(6):871–5.
79. Mussen F, Salek S, Walker S. Benefit-Risk Appraisal of Medicines. A systematic approach to decision-making. 2009.
80. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines – part 1: The development of a new model using multi-criteria decision analysis y. *Pharmacoepidemiol Drug Saf*. 2007;(June):2–15.
81. Nixon RM, Hagan AO, Oakley J, Madan J, Stevens JW, Bansback N, Brennan A. The Rheumatoid Arthritis Drug Development Model: a case study in Bayesian clinical trial simulation. *Pharm Stat*. 2009;(April):371–89.
82. Ohlssen D, Price KL, Xia HA, Hong H, Kerman J, Fu H, Quartey G, Heilmann CR, Ma H, Carlin BP. Guidance on the implementation and reporting of a drug safety Bayesian network meta-analysis. *Pharm Stat*. 2014;13(1):55–70.
83. Ouellet D, Werth J, Parekh N, Feltner D, McCarthy B, Lalonde RL. The use of a clinical utility index to compare insomnia compounds: a quantitative basis for benefit-risk assessment. *Clin Pharmacol Ther*. 2009 Mar;85(3):277–82.

CHAPTER 1

84. Ouellet D. Benefit-risk assessment: the use of clinical utility index. *Expert Opin Drug Saf.* 2010 Mar;9(2):289–300.
85. Pfohl M, Schädlich PK, Dippel F-W, Koltermann KC. Health economic evaluation of insulin glargine vs NPH insulin in intensified conventional therapy for type 1 diabetes in Germany. *J Med Econ.* 2012 Jan;15 Suppl 2:14–27.
86. Phillips LD, Fasolo B, Zafiropoulos N, Eichler H-G, Ehmann F, Jekerle V, Kramarz P, Nicoll A, Lönngren T. Modelling the risk-benefit impact of H1N1 influenza vaccines. *Eur J Public Health.* 2013 Aug;23(4):674–8.
87. Poland B, Hodge FL, Khan A, Clemen RT, Wagner JA, Dykstra K, Krishna R. The clinical utility index as a practical multiattribute approach to drug development decisions. *Clin Pharmacol Ther.* Nature Publishing Group; 2009 Jul;86(1):105–8.
88. Green P, Rao V. Conjoint Measurement for Data Quantifying Judgmental. *J Mark Res.* 2014;8(3):355–63.
89. Sabin T, Matcham J, Bray S, Copas A, Parmar MKB. A Quantitative Process for Enhancing End of Phase 2 Decisions. *Stat Biopharm Res.* 2014 Jan;6(1):67–77.
90. Sadatsafavi M, Marra C, Marra F, Moran O, FitzGerald JM, Lynd L. A quantitative benefit-risk analysis of isoniazid for treatment of latent tuberculosis infection using incremental benefit framework. *Value Health.* 2013;16(1):66–75.
91. Shaffer ML, Watterberg KL. Joint distribution approaches to simultaneously quantifying benefit and risk. *BMC Med Res Methodol.* 2006 Jan;6:48.
92. Shepherd J, Brodin H, Cave C, Waugh N, Price A, Gabbay J. Pegylated interferon a-2a and -2b in combination with raibavirin in the treatment of chronic hepatitis C: a systematic review and economic evaluation. *Health Technol Assess (Rockv).* 2004;8(39).
93. Sherrill B, Amonkar MM, Stein S, Walker M, Geyer C, Cameron D. Q-TWiST analysis of lapatinib combined with capecitabine for the treatment of metastatic breast cancer. *Br J Cancer.* 2008 Sep 2;99(5):711–5.
94. Sherrill B, Wang J, Kotapati S, Chin K. Q-TWiST analysis comparing ipilimumab/dacarbazine vs placebo/dacarbazine for patients with stage III/IV melanoma. *Br J Cancer.* Nature Publishing Group; 2013 Jul 9;109(1):8–13.
95. Smart A. A multi-dimensional model of clinical utility. *Int J Qual Health Care.* 2006 Oct;18(5):377–82.
96. Stojadinovic, Alexander Nissan A, Eberhardt J, Chua T, Pelz J, Esquivel J. Development of a Bayesian Belief Network Model for Personalized Prognostic Risk Assessment in Colon Carcinomatosis. *Am Surg.* 2011;77(2):221–30.

97. Stonebraker JS. How Bayer Makes Decisions to Develop New Drugs. *Interfaces* (Providence). 2002 Dec;32(6):77–90.
98. Thompson JP, Noyes K, Dorsey ER, Schwid SR, Holloway RG. Quantitative risk-benefit analysis of natalizumab. *Neurology*. 2008 Jul 29;71(5):357–64.
99. Tomeczkowski J, Stern S, Müller A, von Heymann C. Potential cost saving of Epoetin alfa in elective hip or knee surgery due to reduction in blood transfusions and their side effects: a discrete-event simulation model. *PLoS One*. 2013 Jan;8(9):e72949.
100. Verduijn M, Grootendorst DC, Dekker FW, Jager KJ, le Cessie S. The analysis of competing events like cause-specific mortality--beware of the Kaplan-Meier method. *Nephrol Dial Transplant*. 2011 Jan;26(1):56–61.
101. Weinstein L, Radano TA, Jack T, Kalina P, Eberhardt JS. Application of multivariate probabilistic (Bayesian) networks to substance use disorder risk stratification and cost estimation. *Perspect Health Inf Manag*. 2009 Jan;6:1-27.
102. Keeney RL, Raiffa H. *Decisions with Multiple Objectives*. Cambridge: Cambridge University Press; 1993.
103. Walker S, Cone M. *The Development of a Model for Benefit-Risk Assessment of Medicines Based on Multi-Criteria Decision Analysis (Summary Report)*. 2004.
104. Breimer D. PK/PD modelling and beyond: impact on drug development. *Pharm Res*. 2008;25:2720–2.
105. Breimer D, Danhof M. Relevance of the application of pharmacokinetic-pharmacodynamic modelling concepts in drug development - the wooden shoe paradigm. *Clin Pharmacokinet*. 1997;32:259–67.
106. Zineh I, Woodcock J. Clinical pharmacology and the catalysis of regulatory science: opportunities for the advancement of drug development and evaluation. *Clin Pharmacol Ther*. 2013;93:515–25.
107. Jonker D, Visser S, van der Graaf P, Voskuyl R, Danhof M. Towards a mechanism-based analysis of pharmacodynamic drug-drug interactions in vivo. *Pharmacol Ther*. 2005;106:1–18.
108. Danhof M, de Jongh J, De Lange E, Della Pasqua O, Ploeger B, Voskuyl R. Mechanism-based pharmacokinetic-pharmacodynamic modeling: biophase distribution, receptor theory, and dynamical systems analysis. *Annu Rev Pharmacol Toxicol*. 2007;47:357–400.
109. Machado S, Miller R, Hu C. A regulatory perspective on pharmacokinetic/pharmacodynamic modelling. *Stat Methods Med Res*. 1999;8:217–45.
110. Sjörgen E, Nyberg J, Magnusson M, Lennernäs H, Hooker A, Bredber U. Optimal Experimental Design for Assessment of Enzyme Kinetics in a Drug Discovery Screening Environment. *Drug Metab Dispos*. 2011;39:858–63.

CHAPTER 1

111. Yousef A, Melhem M, Xue B, Arafat T, Reynolds D, Van Wart S. Population pharmacokinetic analysis of clopidogrel in healthy Jordanian subjects with emphasis optimal sampling strategy. *Biopharm Drug Dispos.* 2013;226:215–26.
112. Lim H, Kim S, Noh Y, Lee B, Jin S, Park H, Kim S, Jang IJ, Kim SE. Exploration of optimal dosing regimens of haloperidol, a D2 Antagonist, via modeling and simulation analysis in a D2 receptor occupancy study. *Pharm Res.* 2013;30:683–93.
113. Sammons H, Starkey E. Ethical issues of clinical trials in children. *Paediatr Child Heal.* 2012;22:47–50.
114. Bellanti F, Della Pasqua O. Modelling and simulation as research tools in paediatric drug development. *Eur J Clin Pharmacol.* 2011;67(Suppl 1):75–86.
115. Cella M, Zhao W, Jacqz-Aigrain E, Burger D, Danhof M, Della Pasqua O. Paediatric drug development: are population models predictive of pharmacokinetics across paediatric populations? *Br J Clin Pharmacol.* 2011;72:454–64.
116. Manolis E, Osman T, Herold R, Koenig F, Tomasi P, Vamvakas S, Saint Raymond A. Role of modeling and simulation in pediatric investigation plans. *Paediatr Anaesth.* 2011;21:214–21.
117. Manolis E, Pons G. Proposals for model-based paediatric medicinal development within the current European Union regulatory framework. *Br J Clin Pharmacol.* 2009;68:493–501.
118. EMA. Benefit-risk methodology project - Report on risk perception study module. 2012.
119. Tervonen T, Van Valkenhoef G, Buskens E, Hillege HL, Postmus D. A stochastic multi-criteria model for evidence-based decision making in drug benefit-risk analysis. *Stat Med.* 2011;30(12):1419-28
120. Phillips L. Improving the process of balancing benefits and risks in approving drugs (presentation). EMA Decision Analysis Affinity Group Workshop; 2010.
121. Phillips L. Benefit-Risk modelling of pharmaceuticals: Where are we now? (presentation). EFSPi-FMS-DSBS Benefit Risk Assessment Methodology Workshop 7; 2012.
122. Lam G, Petri M. Assessment of systemic lupus erythematosus. *Clin Exp Rheumatol.* 2005;23:S120–S132.
123. French S, Xu D. Comparison study of multi attribute decision analytic software. *J Multi Criteria Decis Anal.* 2005;13:65–80.
124. EMA. EPAR Benlysta: Summary of product characteristics. 2013.
125. Bender BC, Schindler E, Friberg LE. Population pharmacokinetic pharmacodynamic modelling in oncology: a tool for predicting clinical response. *Br J Clin Pharmacol.* 2013;79(1):56–71.

126. Pink J, Lane S, Hughes DA. Mechanism-Based Approach to the Economic Evaluation of Pharmaceuticals. *Pharmacoeconomics*. 2012;30(5):413–29.
127. Naci H, van Valkenhoef G, Higgins JPT, Fleurence R, Ades AE. Evidence-Based Prescribing: Combining Network Meta-Analysis With Multicriteria Decision Analysis to Choose Among Multiple Drugs. *Circ Cardiovasc Qual Outcomes*. 2014;7:787–92.
128. Van Valkenhoef G, Tervonen T, Zhao J, De Brock B, Hillege HL, Postmus D. Multicriteria benefit-risk assessment using network meta-analysis. *J Clin Epidemiol*. Elsevier Inc; 2012;65(4):394–403.
129. Chain ASY, Dieleman JP, van Noord, Charlotte Hofman A, Stricker BHC, Danhof M, Sturkenboom MCJM, Della Pasqua, O. Not-in-trial simulation I: Bridging cardiovascular risk from clinical trials to real-life conditions. *Br J Clin Pharmacol*. 2013;76(6):964–72.

Appendix

Decision tree

The theory underlying the use of decision trees is based on visualisation of the decision making process by a branching structure with decisions as roots, and possible outcomes as branches. In this way, decisions, subsequent uncertain events, consequences and multiple criteria are described (14). Complexity arises with addition of extra nodes. In the EMA framework for benefit-risk assessment, this technique has been used as a link between qualitative and quantitative approaches, as it encompasses objectives and possible outcomes, combined with numerical data on frequencies and uncertainties, with which benefit-risk balance can be calculated for each outcome or decision. Figure S1 shows an example of a decision tree. The decision (square node) consists of two alternatives: approving the vaccine against H1N1 swine flu by the end of September, or waiting until October, so more data can be gathered on efficacy and safety. The remaining uncertainties are modelled as events (round nodes), for which consequences the working group determined the probability, mostly based on earlier experience. Disease seriousness for example has a probability of 20% to become severe, based on the historical observation that one in five pandemics becomes catastrophic (Spanish flu). For the delay, this probability increases, as early vaccination can prevent escalation. All other probabilities are determined and the estimated deaths or serious disabled (DSDs) are stated for all 24 outcomes (triangle nodes). The decision tree itself enables back calculation to the actual decision, providing the consequence of each alternative would be. This is achieved by multiplying outcomes with the probabilities as weight. Taking the best case scenario, the working group determined that in the case of early approval, moderate disease seriousness probability is 0.8, probability of efficacy of 75% is 0.3, with probabilities of safety events rate of 1/100,000 and 1/10,000 at 0.9 and 0.1, respectively. The latter outcomes are stated to be 42500 and 87500 DSDs, so calculating back, the average DSD-value of the vaccine in early approval in case of a moderate disease with an efficacy of 75% is 47000 DSDs. Applying these steps for all outcomes and events results in an average DSDs for the two alternative options, in this case 216,500 for September and 291,547 for October, showing that earlier approval is the better option, a decision made after completion of a sensitivity analysis of the chosen probabilities (86). A reconstruction of the complete decision tree can be found in (120). This concept alone is valuable when cases remain relatively simple. Although the tree itself might be too complex in advanced cases, it remains an important building block for more evolved techniques, such as MCDA.

MCDA and the EMA's BR framework

When evaluating very complex scenarios with multiple endpoints and objectives, it becomes crucial to have a clear understanding of the context structure. In very simple words, MCDA allows breaking up the problem and analyzing individual factors, before reassembling each component to provide a thorough overview of the analysis and making a final decision (102). In this structure, it combines the decision tree theory with value functions. In other words, it converts the different inputs for the decision model into preference values, allowing comparing the different endpoints on a common ground. The preference value scale requires probability, utility and the preference of the alternative associated with the highest expected utility. Multiple objectives are evaluated together on different identified criteria and a balance is made after scoring and weighting these criteria, with uncertainty taken into account (14). As highlighted by the EMA BR project, the use of the ProACT-URL approach in combination with decision tree and MCDA represents a more transparent and consistent assessment of the BR balance (44). The technique consists of eight steps that will be briefly discussed in the following paragraphs. In the next section, these steps are illustrated using Benlysta (belimumab), a drug against Systemic Lupus Erythematosus (SLE), as a paradigm compound (121).

Step 1 to 3: defining decision context; identifying options, eg. study arms; identifying objectives and criteria, eg. maximising benefit and minimising risk, more specified in a decision tree. This part of MCDA overlaps with the ProACT-URL approach: creating a qualitative framework of objectives and context, as well as with the decision tree, as mentioned earlier (120). Benlysta has been proposed for the treatment of adults with high disease activity, with autoantibody-positive SLE. It should be added to the standard treatment, which consists of hydroxychloroquine and corticosteroids (step 1). The available studies include two randomised, placebo-controlled clinical trials and three open-label continuation safety trials. The dosing regimens used in these trials include either 1 or 10 mg (step 2). There is a medical need for newer, more effective and better tolerated therapies. To specify these criteria, an effect tree is composed, as visualised in Figure S2 (step 3).

Step 4 and 5: scoring; weighting. Scoring and weighting are the most important steps as their aim is to normalise the raw input data for the decision model by translating them into preference values. Scoring means scaling each criterion (input data characterised by different units and time scales) by assigning a new range, which is usually set between 0 and 100. Within this range, different outcomes are directly or indirectly scored, where the ratio of difference is the most important. Scoring for Benlysta, as visualised in Table S2, is performed following defined clinical scales, like SLE Disease Activity Index (SLEDAI). First, the two extremes are evaluated, best and worst with corresponding units, after which the three options are considered within this range.

The weighting step normalise all measures into one preference scale, judging which criteria is more important and allowing comparing the different options into one common level (103). This procedure allows translating the scoring into preference values, which carries the subjective component. Weighting can be done linear, direct or inverse, or non-linear. Finally, swing weights are assigned, based on trade-offs among favourable or unfavourable effects, or between the most important favourable and unfavourable effects. In other words, if objective A is twice as important as objective B, the score doubles on that scale. These swing weights depend on the subjective choice considering the relative difference in original scale and the importance of the corresponding objective to the whole. Considering as an example buying a car, limiting costs is an objective of importance. If, however the difference between alternatives in this criterion is only small, the impact of that objective becomes limited (79). It is also important to take into account the possibility of single events that are multiple times considered. In this specific case, the SLE assessment scores SLEDAI and BILAG-index have similar criteria, like psychosis or vasculitis. If this is not corrected by the assignments of weights, these events have double impact on in this case the unfavourable effects (122).

Step 6 and 7: combining data to overall value; examining results. The overall score is simply the sum of the product of the score and weight per criterion, as stated in equation 1, where S_i is the overall score per option i on criterion j , with s_{ij} as preference score of the option and w_i as the weight of the criterion (41).

$$S_i = w_1s_{i1} + w_2s_{i2} + \dots + w_n s_{in} = \sum_{j=1}^n w_j s_{ij} \quad \text{eq.1}$$

This aggregation is performed by software; several are currently available for this methodology (e.g., HiView, V.I.S.A., Web-Hipre, Expert Choice, Logical Decisions) (123). Cumulative weights are calculated based on the normalized weight; overall weighted scores per options are visualized graphically (Figure S3).

Step 8: perform sensitivity analysis. The sensitivity analysis is important to identify possible judgments of serious impact, thus reducing uncertainty. Displaying the variation of weights on each criterion allows identifying possible crossovers at which a change in the relationship between weight and criterion might be observed for the different options.

Table S1. Overview of quantitative methodologies to assess benefit-risk balance, as given by the CHMP [14].

Method	Advantages	Limitations	References
Bayesian beliefs networks	Network of nodes representing risks, benefits, observations and assessments, connected by conditional arrows, which input probabilities result in probability distribution for all nodes. Inclusion of both objective data and subjective expert opinion. Visualisation of effect of factors on each other.	Requires structural similarity across cases, which in BR might only be appropriate for similar indications. Probability input as a subjective element remains unsupported. Uncertainty of indirect effects introduces bias in their impact on the outcome.	(14,29,82,96,101)
Bayesian statistics	Prior and posterior probabilities based on available evidence. Together describe the likelihood of an effect and its uncertainty, combined with utility function in the Bayesian approach. Methodology improves as more data are gathered, as it involves iterative learning.	Significance levels state something about data, not hypotheses, so cannot directly be included into a formal BR assessment. The model itself doesn't include multiple criteria. Mathematical models can get complex.	(14,27,53,64,81,89)
Clinical Utility Index	Multi-attribute utility analysis with weighted trade-offs. Utility function introduces clinical meaning to the assessment. CUI is flexible over different indications and endpoints. Transparent method with possibility of sensitivity analysis.	In case of limited applicable data, complex modelling with high variability and uncertainty is required. Subjective discussion on clinically relevant factors remains unsupported. More useful for a no-go than for a go-decision	(45,56,62,65,68,83,84,87,95)
Conjoint analysis	Covers preferences of different stakeholders, utility weight is based	Labour intensive if all stakeholders are included. Weight might not be	(14,54,60,88)

	on preferred trade-offs. Realistic method helpful in weighting.	independent from methodological decisions. Does not account for uncertainty.	
Contingent valuation	Benefits are translated to financial values by enquiring the prize patients are willing to pay for it.	Not focused on BR assessment	(14)
Decision tree	Overview of all possible outcomes with their probabilities, calculated using the branches and nodes leading to said outcome. The decision tree is a useful framework.	Too simple for complex cases. Uncertainties are only limited covered, as probabilities are often empirically determined.	(14,30,74,80,86,97,124)
Discrete event simulation	Detailed simulation based on differential equations and continuous variables. Ability to handle multiple assumed characteristics and simultaneously assess impact of multiple effects on health economics.	Complexity, complicate adaptability, lack of transparency and validation. Risk of underestimation in case of prediction limited to short term effects. No clear assessment of unfavourable effects.	(14,28,31,51,66,71,85,99)
Evidence-based BR model	Model visualised as a set of scales, including the benefit 'box' with efficacy, including responder rate and evidence and the risk 'boxes', for each ADE, with seriousness, frequency and evidence. The method correlate to EMA's definition, as the first two criteria of either box are (un)favourable effects and the third includes uncertainty of effects.	Simplified multi-criteria model with limited (three) criteria each. There is no application supporting the translation of effects into one unit. Preference weights are not accounted for.	(14,19)
Incremental net	Incremental net health benefit is the	Although this is a version of a multi-	(1,14,23,32,33,72,90)

health benefit	difference between unfavourable effects and favourable effects derived from the treatment options, where all effects are normalised into one unit. The method is transparent and theoretically sound, including uncertainty and extrapolation in time.	criteria model, such as MCDA, translation to a single unit requires another methodology (e.g., value-adjusted life years, or QALY). QALY can only be transferred to health benefit when costs are not considered, in other words the willingness to pay is infinite. Weighting of effects is also dependent on another methodology, like conjoint analysis. This methodology on itself is incomplete. Subject to bias by confounders.	
Kaplan-Meier estimation	Function of survival over time, impact measured in ratio of differences, useful in Markov models.	Limited representation of (un)favourable effects, for example in non-fatal indications. It does not account for uncertainty and cumulative probabilities can be misleading due to lack of correlation structure (e.g., competing events).	(14,49,100)
Markov model	Describes time-dependent dynamic processes, using transitions between health states and their probability distributions.	Probability data might be sparse before approval. Complex health states might be oversimplified.	(14,42,43,52,98)
Minimum Clinical Efficacy	The method allows incorporating risks and benefits into one single metric. In addition, relative utilities can be considered during the analysis.	The statistical properties are not yet fully understood and the methodology does not allow characterising the uncertainty around the benefit-risk measurements.	(1,40,57,58)
MCDA	Multi-criteria method breaking up	Might be too comprehensive for	(1,12,14,22–

	the problem, followed by scoring and weighted assessment of benefits and risks as most representative presentation of data. Sensitivity analysis prevents unwanted impact. Incorporates uncertainty.	simple analysis. Does not account for possible correlations between endpoints. Preference value determination is accounted for in the weighting step.	24,26,34,44,47,63,79,80,102,103)
NNT	Easy understandable measure used in the clinic, stating the number of patients required to treat one occurrence of the disease (or to have one more ADE in NNH). Patient preferences can be included using Relative Value Adjusted NNT (RV-NNT).	Limited statistical power and because of lack of preference data, misinterpretation by different risk perceptions, as well as by using the same scale without proper weighting effects. Ratio of NNT/NNH assumes independence and similar timescale.	(1,4,14,22,23,36,46,48,57,58,67,69)
Principle of threes	Simplified method in which only three criteria per risk/benefit are scaled with three possible outcomes (e.g., low, medium, high), benefit and risk are summed up.	Very limited in number of criteria. No weighting of the criteria.	(14,17,35,76)
Probabilistic simulation	Complementary to point estimate statistics, as it states the impact of risk and benefit as a probability distributions based on simulated random draws from study data. More precise, accounts for uncertainty in trade-offs. Can account for correlation, if suitable data is available.	Limited if using non-validated or non-representative probability distributions for simulation. Benefits or risks are not weighted, shown by the fatal adverse event in the adalimumab-study, which did not seem to affect the simulation analysis.	(1,14,23,70,72,73,78,91)
QALY	Multiple dimensions are scored and	Limited in uncertainty and unique	(14,25,92)

INTEGRATION OF PKPD RELATIONSHIPS INTO BENEFIT-RISK ANALYSIS

	weighted for preference, outcome measured in life years on population level.	(disease/patient) data representation, more focussed on health- and pharmacoeconomics. Threshold is debatable.	
Q-TWiST	The method is used to convert time into QALYs; time lost due to an ADR is subtracted to time gained from receiving the treatment. Q-TWiST allows comparing benefits and risks into a single metrics. Furthermore, allowing the inclusion of patients' preferences is considered a valid tool for individual BR assessment.	Although valid for individual assessments, it gives more difficulties to evaluate BR on a population level. Does not allow measuring uncertainty around QALYs. The data needed for the analysis might be difficult to acquire. In addition QALYs might have a major influence on the BR outcome.	(1,22,37,39,49,50,59,75,77,93,94)
Stated preferences	Collection of methods using preference values to determine utility functions of different stakeholders. Measures e.g., the extent patients are willing to experience unfavourable effects to achieve favourable effects.	Empirical method that does not account for uncertainty or weighting. Overlaps with conjoint analysis. Gathering of individual patient data is time consuming.	(1,14,23,38)
System dynamics	Account for non-linearity using feedback and time-delays, both short and long term. Possibility of input data from different sources.	No recorded use in drug development. Focus on pharmacoeconomics. No consideration of (weighted) unfavourable effects, such as ADEs.	(14,55)
TURBO	Simplified method in which only two criteria per risk/benefit are scaled up with five possible outcomes. Pairs of outcomes are weighted and assessed. Frequency, probability,	Very limited in terms of the number of criteria. There is no way of knowing prior to assessment which criteria to choose. Choice might be arbitrary. No theoretical basis.	(14,23)

	severity and extent are included into the choices of criteria.		
--	--	--	--

Table S2. Scoring of Benlysta according to the different criteria, as visualised in the decision tree. FE and UFE are favourable and unfavourable effects, respectively. SRI is SLE Response Index, SLEDAI is SLE Disease Activity Index, PGA is Physician’s Global Assessment, BILAG is British Isles Lupus Assessment Group, where A indicates severe disease and B less active disease. Secondary favourable endpoints are CS, corticosteroids, Flare rate meaning number of new BILAG A cases and QoL measured as mean change in the total score of Short Form 36. SAE are serious adverse events, such as tumour development, opportunistic infections or progressive multifocal leukoencephalopathy (PML). * > 25% and to less than 7.5 mg/day, ** per patient year. Based on (15).

Effects		Name	Description	Best	Worst	Units	Placebo	10 mg	1 mg
Favourable	SRI	SLEDAI	Improved ≥ 4	100	0	%	41	53	48
		PGA	No worsening	100	0	%	66	75	76
		PGA	Mean change	1	0	Difference	0,44	0,48	0,45
		BILAG	No new A/2B	100	0	%	69	75,2	70,1
	Secondary endpoints	CS sparing	Dose reduction*	100	0	%	12,3	17,5	20
		Flare rate	New BILAG A cases**	0	5	Frequency	3,51	2,88	2,9
		QoL	Mean change SF36	0	100	Difference	3,5	3,4	3,7
Unfavourable		SAE	Potential	100	0		100	0	90
		Infections	Life-threatening infections	0	10	%	5,2	5,2	6,8
		Sensitivity reaction	Hypersensitivity reactions	0	2	%	0,1	0,4	0,3

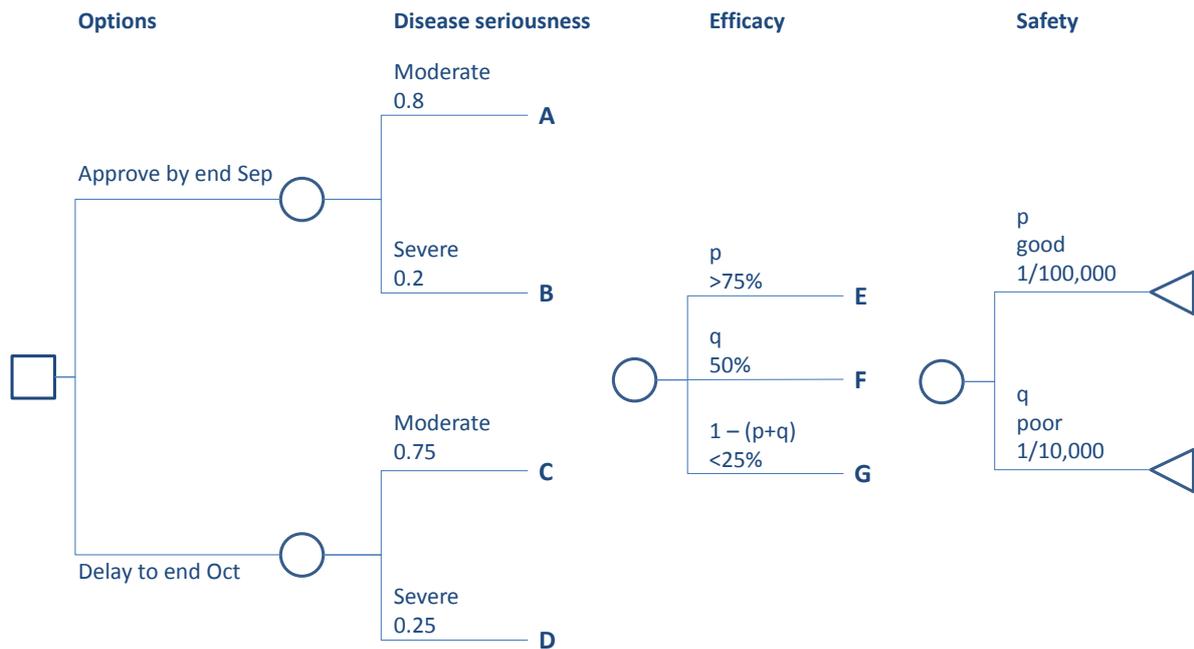


Figure S1. Example of decision tree concerning approval of swine flu vaccine in 2009, where decision of approval planning is followed by the consequences for disease seriousness. Efficacy branches attach to A through D, whereas safety branches to E through G resulting in 24 scenarios with calculable event outcomes. Based on (116).

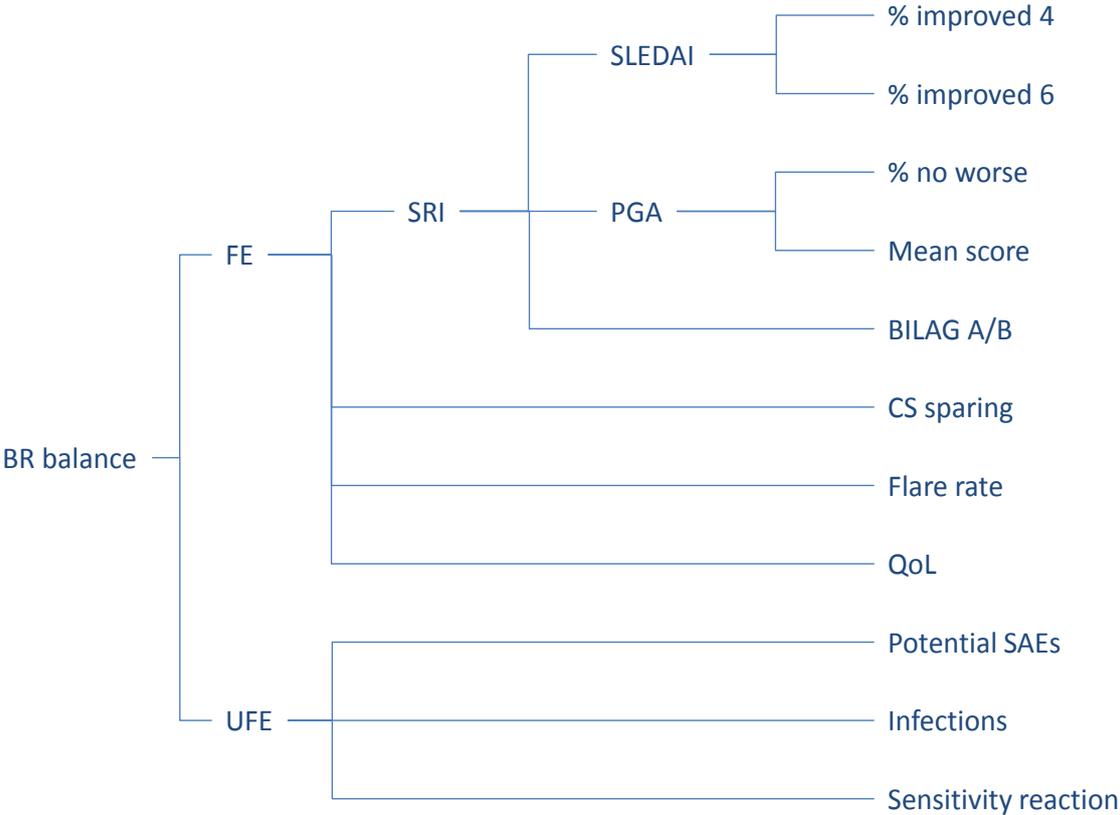


Figure S2. Outcomes tree based on identified criteria used for the Benlysta example. FE and UFE are favourable and unfavourable effects, respectively. SRI is SLE Response Index, SLEDAI is SLE Disease Activity Index, PGA is Physician’s Global Assessment, BILAG is British Isles Lupus Assessment Group, where A indicates severe disease and B less active disease. Secondary favourable endpoints are CS, corticosteroids, Flare rate meaning number of new BILAG A cases and QoL measured as mean change in the total score of Short Form 36. SAE are serious adverse events, such as tumour development, opportunistic infections or progressive multifocal leukoencephalopathy (PML). Based on (15).

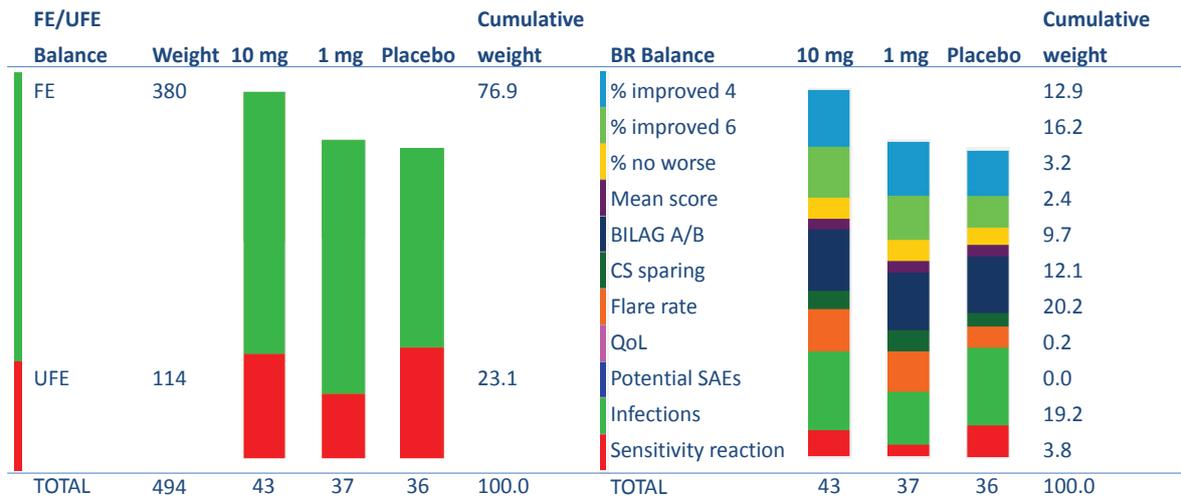


Figure S3. Different data presentations to evaluate benefit risk balance, in which the cumulative weight is calculated and the overall weighted scores are visualised. Left, impact of favourable effects (FE) and unfavourable effects (UFE) are shown in green and red, respectively. On the right, all different criteria are shown with their impact, which results in a more informative presentation of the data. For example, sensitivity reaction to 1mg has decreased impact as compared to placebo or 10mg. Based on (15).

CHAPTER 2

Model-informed benefit-risk assessment of iron chelation in transfusion-dependent haemoglobinopathies

Scope and intent of the investigation

2.1 General introduction

Drug approval by regulatory agencies is granted on the basis of the evidence on the safety and efficacy profile of a drug, which has been generated throughout the drug development phases (1–4). At this stage, decisions about the benefit-risk balance rely on the assumption that the data collected are sufficient to allow an unbiased evaluation of the safety and efficacy of a given intervention. This assumption may not be valid for all drugs, with a vast number of conditions and diseases in which numerous clinical questions cannot be fully addressed at the time of approval. In fact, the use of a question-based approach for the review of regulatory submissions by some regulatory agencies has highlighted the relevance of understanding which clinical and scientific questions need to be considered for the approval of a new drug. To be effective, such a regulatory process requires sponsors, researchers and clinicians to reflect on which data need to be generated, what is already known and how both existing and new data are integrated and processed. Moreover, as widely recognised by different stakeholders, including regulatory authorities, industry and patients (5–7), a clear framework for the assessment of benefit-risk balance (BRB) in which quantitative methods are used to translate findings obtained during the development process into measures that summarise favourable and unfavourable effects of treatment, is still lacking. This situation has resulted in undefined, inconsistent and non-transparent decision making (8–10).

Clarity about which clinical and scientific questions need to be addressed as well as the availability of quantitative methods to translate findings into summaries of favourable and unfavourable effects are requirements that apply to all drugs, but they become even more relevant when dealing with special populations, such as the paediatric population, where practical and ethical constraints make the process of generating evidence extremely

challenging (11–14). These hurdles limit the level of evidence available at the time of the first-marketing authorisation, as compared to other populations. In this thesis we focus on the current challenges in evidence generation during paediatric drug development. We demonstrate how modelling and simulation (M&S) can be applied for evidence synthesis and decision making; by the integration of existing and new data, to address essential clinical and scientific questions and to support a more comprehensive evaluation of the benefit-risk balance of a medicinal product prior to its approval. In addition, the implementation of such a framework will allow for better understanding of consequences of an intervention, and consequently improve risk management and therapeutic use of medicinal products. The examples provided in the following chapters were developed in the context of paediatric diseases, but the concepts underpinning the proposed framework can be extrapolated to a broader range of diseases and conditions across any patient population.

In order to demonstrate the contribution of modelling and simulation as a tool for more effective data generation, evidence synthesis and better decision making, the work presented in this thesis will be divided into three main sections, namely:

1. Optimisation of study protocol design and data generation in children ;
2. Integration of existing knowledge and mechanism-based parameterisation of drug- and disease-specific properties;
3. Use of clinical trial and not-in-trial simulations to complement data generation and improve benefit-risk assessment.

An outline of the scope of the research and details on the implementation of the different sections are presented in the next paragraphs.

In **chapter 1**, an overview of the different methodologies currently available for benefit-risk assessment (BRA) is presented. Focus is given to differences between qualitative and quantitative approaches and the relevance of the latter for accurate decision making about the benefit-risk profile of a medicinal product. As recently suggested by EMA, amongst the available approaches the use of Multi Criteria Decision Analysis (MCDA) appears to have the right features to address the lack of transparency in the way benefit-risk balance is assessed, enabling integration of different dimensions or levels of clinical concern during the process (15).

Our review highlights how MCDA can benefit from the use of M&S in order to better define the BR balance of a given drug and vice versa, i.e. how concentration-effect relationships can provide a stronger basis for understanding benefit and risk, and how pharmacologists can gain insight into the therapeutic value of an intervention by jointly evaluating multiple

endpoints. The advantages of such an integrated approach are illustrated by the few available examples in the published literature.

The theoretical concepts presented in chapter 1 form the basis for the experimental work proposed in this thesis, which will be described in the subsequent paragraphs in this chapter. We will make use of chronic iron overload by transfusion-dependent haemoglobinopathies as a case study. Iron overload provides all the necessary elements, i.e., a complex multidimensional disease condition with short- and long-term complications that can lead to different clinical presentations over time. Moreover, it has a sufficiently low incidence to allow lessons learned to be applied in other rare paediatric diseases.

2.2 Transfusion-dependent haemoglobinopathies

Among the transfusion-dependent diseases, β -thalassaemia major is one of the most common disorders. It belongs to a group of hereditary blood disorders characterised by reduced or absent beta-globin chain synthesis. As a result, patients suffer from reduced haemoglobin (Hb) levels in red blood cells (RBC) and decreased RBC production followed by anaemia (16).

Historically, the majority of thalassaemia patients are located in the Mediterranean countries, in the Middle East and Asia. According to the Thalassaemia International Federation (TIF), around the world only about 200.000 patients are alive and registered as receiving regular treatment (17). Children are usually diagnosed between 6 and 24 months after birth. Early clinical symptoms include feeding problems, diarrhoea and progressive enlargement of the abdomen caused by spleen and liver enlargement. In some developing countries, patients also suffer from growth retardation, poor musculature and skeletal changes (18). Individuals affected by β -thalassaemia major require regular RBC transfusions to survive. Without transfusions or in the presence of poor management of the disease, patients often die before the third decade of life. According to the guidelines for the clinical management of thalassaemia (17) transfusion intervals should aim to maintain a pre-transfusion Hb level between 9 and 10 g/dl and a post-transfusion level of 14 to 15 g/dl. This requirement implies the need for frequent blood transfusions, with the most common transfusion interval being once every two to four weeks (equal to two to three blood units per three weeks).

A graphical overview of iron distribution and storage is provided in figure 1. In the context of transfusion-dependent diseases, it worth mentioning that there is no innate mechanism that is able to clear any iron excess from the body. Under normal physiological conditions, iron is almost completely recycled within the body.

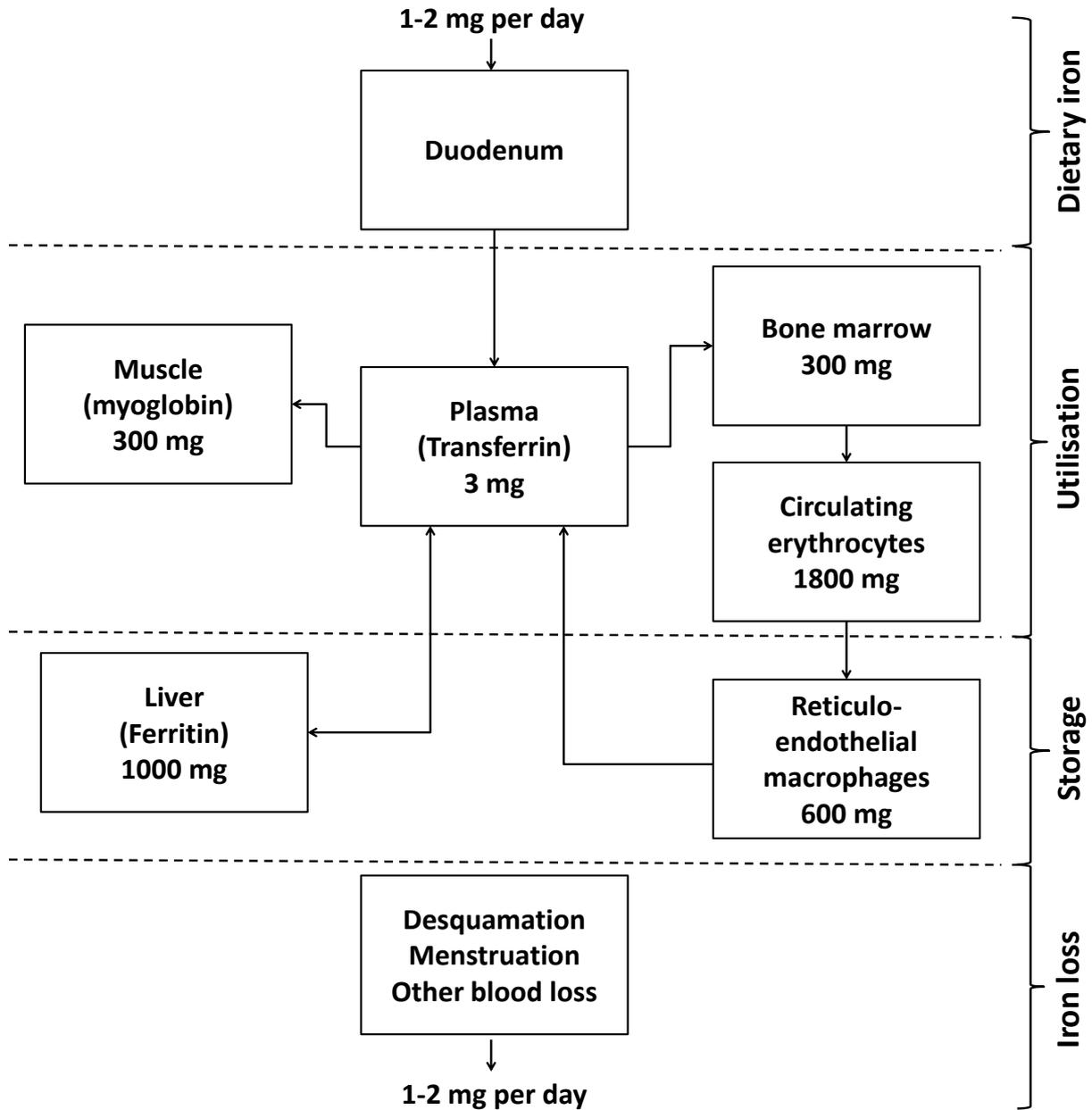


Figure 1. Iron homeostasis. In a balanced state, 1 to 2 mg of iron enters and leaves the body each day. Dietary iron is absorbed by duodenal enterocytes. It circulates in plasma bound to transferrin. Most of the iron in the body is incorporated into haemoglobin in erythroid precursors and mature red cells. Approximately 10 to 15 percent is present in muscle fibres (in myoglobin) and other tissues (in enzymes and cytochromes). Iron is stored in parenchymal cells of the liver and reticuloendothelial macrophages. These macrophages provide most of the usable iron by degrading haemoglobin in senescent erythrocytes and reloading ferric iron onto transferrin for delivery to cells. Adapted from: Andrews et al, *N Engl J Med.* 1999 (19).

Iron entry into the cells is regulated by the uptake of iron-transport protein transferrin from the plasma. Once chronic RBC transfusion therapy has started, iron exposure in macrophages increases, which results in the saturation of transferrin transport capacity. This leads to the release of non-transferrin bound iron (NTBI) in plasma. NBTI can then enter important tissues (e.g., in heart and liver) and accumulate over time. As iron is stored in tissues mainly as ferritin complexes, once ferritin storage capacity has saturated small clusters of ferritin particles will be formed and degraded by lysosomes leading to the formation of insoluble masses of hemosiderin (20–26). Over time these masses can cause severe organ damage (19,27–31).

Iron overload and chelation therapy

Even though significant improvements have been achieved in the management of the chronic transfusion regimens in the past decades, RBC therapy will eventually lead to a series of complications. Iron overload is the most common and relevant one and it is associated with several (lethal) co-morbidities such as cardiac dysfunction, liver fibrosis, hypogonadism, hypothyroidism, hypoparathyroidism and diabetes mellitus (28,30). Cardiac disease caused by myocardial siderosis is the most relevant complication, causing death in 71% of the patients affected by transfusion-dependent diseases (27). In the absence of an innate mechanism that allows removing iron excess from the body, treatment with iron chelators is essential to prevent iron accumulation and related complications (32–35). An overview of iron chelators currently approved for the treatment of iron overload is shown in table 1 (33).

Table 1. Summary of the available iron chelators. Adapted with permission from: Kwiatkowski JL. *Pediatr Clin N Am.* 2008; 55:461-82 (33)

Property	Deferoxamine	Deferiprone	Deferasirox
Chelator:iron binding	1:1	3:1	2:1
Route of administration	Subcutaneous or intravenous	Oral	Oral
Usual dosage	25-50 mg/kg per day	75 mg/kg per day	20-30 mg/kg per day
Schedule	Administered over 8-24 hours, 5-7 days per Week	Three times a day	Daily
Adverse events	Local reactions Ophthalmologic Auditory Pulmonary Neurologic Infectious	Agranuloctyosis/neutropenia Arthralgias/Arthritis	Gastrointestinal disturbances Renal Insufficiency
Advantages	Long-term data Available	May be superior in removal of cardiac iron	Once daily administration Only oral chelator licensed for use in US
Disadvantages	Toxicity Compliance Problems	Not licensed for use in United States. Frequent blood count monitoring required	Long-term data lacking
Drug cost	\$	\$\$	\$\$\$

Iron chelators possess a similar mechanism of action. They act by 1) preventing the uptake of NTBI into organs, such as liver and heart; 2) chelating intracellular iron and thus preventing its incorporation into ferritin; or 3) intercepting iron released from degraded ferritin (36).

Clinical assessment of iron overload

The symptoms and signs associated with iron overload can be initially diagnosed and assessed by different clinical biochemistry parameters. The most common marker of iron imbalance is serum ferritin, which indirectly reflects the correlation between circulating levels and total body iron stores (37). The use of serum ferritin alone, as a single clinical marker however is not always sufficiently robust to detect iron overload. Ferritin levels are also be influenced by other factors such as inflammatory disorders and liver disease (38).

Therefore, serial measurements of serum ferritin are still the easiest and least invasive method to evaluate iron overload and efficacy of chelation therapy.

Other methods for the assessment of iron overload focus more on tissue specific accumulation. Liver iron concentration (LIC) is considered as the gold standard for the evaluation of iron overload. LIC has been shown to correlate well with total body iron accumulation (39). The measurement of liver iron concentration requires, however, an invasive technique, which may lead to potential clinical complications and bias, such as in the case of false negative results (40). Magnetic bio-susceptometry (SQUID) is another option for measurement of liver iron accumulation (41). However, it is only available in a limited number of centres worldwide. Furthermore, cardiac complications due to iron accumulation in the heart have been associated with 50-70% of deaths in thalassaemia major patients, mainly at young age (42). Methods for cardiac monitoring were developed under the assumption that keeping serum ferritin and LIC level below a certain threshold (<2500 µg/L and <7 mg/kg dwt respectively) would lead to decreased cardiac risks. However, cardiac dysfunctions were often identified at relative late stage of treatment, suggesting that this method was not sufficient for effective intervention. In recent years, magnetic resonance imaging (MRI) techniques for assessing iron loading in the liver and heart have been introduced and validated for the evaluation of tissue specific accumulation (43).

Clearly, understanding and integration of knowledge about the short and long term mechanisms underlying iron overload are lacking. The ability to predict iron organ accumulation based on systemic, non-invasive markers such as ferritin will depend on further characterisation of dynamic, homeostatic processes. In this context, accurate details of the transfusion history and assessment of the effects of chelation therapy are equally important.

In the next sections we present details of the investigations, which will provide the basis not only for the characterisation of the disease, but also for the design and optimisation of clinical protocols in children. These concepts are followed by the introduction to methods supporting evidence synthesis as a means to better understand the safety and efficacy profile of a drug. Two important aspects reflect the novelty in the approach described here. First the use of a multi-model analysis in which different measures of efficacy and safety are derived according to underlying biological or pharmacological correlations, where applicable. Second, the integration of clinical data from real and virtual patients, whose responses are simulated from the aforementioned models, to improve the assessment of benefit-risk profile by multi-criteria decision analysis (MCDA).

2.3 Optimising evidence generation in paediatric trials

Practical and ethical constraints to the implementation of clinical trials in the paediatric population (11,12,14), make evidence generation in most paediatric diseases extremely challenging. The value of the new data is tremendously higher than in a standard protocol involving adults. Yet, little attention has been paid to the opportunities to ensure that high quality data are obtained whilst keeping the burden for the children to a minimum.

As indicated previously, the approval of a medicinal product relies on the ability of a sponsor to address clinical and scientific questions regarding the efficacy and safety profile of the drug under investigation. Here factors such as how knowledge is generated in this population and which type of data is needed to approve a given therapeutic intervention ultimately underpin the validity of the experimental evidence provided in a regulatory submission. In a very simplistic manner, it can be said that three scenarios have been used to determine the rationale for paediatric programs, while relying on adults as a reference population: 1) if the disease has different features in adults and children, then both pharmacokinetic and efficacy/safety data must be generated; 2) if the disease and its progression as well as the main endpoints of interest are similar in the two populations bridging concepts can be applied and pharmacokinetic and eventually pharmacodynamic data should be sufficient to prove comparable efficacy; 3) in some cases it is also conceivable that pathophysiological processes and pharmacological mechanisms are sufficiently understood to allow extrapolation of efficacy findings from the adult population without the need of generating new evidence in children. In all three cases the quality of the data collected is crucial to establish not only the effect size of a treatment, but also to define the actual benefit-risk profile of the intervention.

From a clinical and scientific point of view, this implies that high accuracy and precision are desirable, irrespective of the nature of the trial.

Whilst the aforementioned scenarios are valuable steps to mitigate the burden of evidence generation in children, they also imply the need for generating evidence prior to approval as a key requirement. None of these scenarios formally considers how current understanding of a drug, disease or patient population in adults can contribute to the decision making process for children. We foresee the integration of available knowledge with clinical data can significantly improve one's ability to assess the benefit-risk profile of a treatment and reduce the uncertainty associated with gaps in the data available at the time of submission. From a clinical pharmacology perspective, this implies that the concept of bridging could be expanded to situations where the disease is different in children and adults. If, such differences are simply due to the natural course of the disease, then these differences may be predicted by parametric (mathematical) representation of the underlying processes in a drug-disease model. This is the situation that we deal with throughout this thesis, i.e.,

haemoglobinopathies, in which long-term complications, which are the primary consequence of iron overload, clearly mark the difference between adults and children and can be anticipated using prior knowledge.

Amongst the opportunities for increasing the informative value of data collected in children is the possibility of using population pharmacokinetic or pharmacokinetic-pharmacodynamic modelling in conjunction with optimal design concepts to reduce sample size and frequency in the so-called bridging studies.

Despite the wide clinical experience with iron chelators, and more specifically with deferiprone, there is no pharmacokinetic data in children below 6 years of age. Given the nature of the disease and its progression, a model-based approach can be used to optimise a prospective pharmacokinetic study in children and consequently define the dosing requirements in this subgroup. First, we demonstrate in **chapter 3** how available pharmacokinetic data from adults and adolescents can be characterised by means of a population pharmacokinetic model. We then explore how uncertainty about the changes in pharmacokinetic properties of deferiprone can be evaluated in conjunction with optimal design theory. A proposal for sampling schedule and group size is presented in **chapter 4**, where ED-optimality concepts are used to identify the most suitable sampling scheme in the absence of data in the population under investigation. This information will be used to support the design of a prospective bridging study in children with less than 6 years of age. Subsequently, we show how modelling and simulation enables the evaluation of the pharmacokinetics of deferiprone based on sparse data. Dosing recommendations are proposed based on the predicted exposure to deferiprone taking into account the parameter distributions in the target patient population. In this investigation, it is worth mentioning that dosing recommendations involve more than simply the data obtained in a small group of children: it encompasses parameter-covariate interactions, which may not be well represented in the trial population.

Whereas these concepts have been implemented for a specific drug, a similar approach can be applied for the evaluation of biomarkers or clinical response.

2.4 Integrated evaluation of efficacy and safety by modelling and simulation

The concepts underpinning the optimisation of pharmacokinetic data collected in prospective clinical trials are also extremely important in the evaluation of the pharmacodynamics of a drug. In this sense, population PKPD modelling can be applied as a tool for evidence synthesis. In addition to the opportunities to increase the informative value of data collection in prospective studies, PKPD modelling also address a critical aspect

of clinical pharmacology research, i.e., the integration of information as the basis for the evaluation of treatment response when complex and multiple factors are involved. Moreover, the approach enables one to account for multidimensionality, i.e., to evaluate in an integrated manner multiple endpoints. The correlation or interdependency between endpoints or measures of drug response is currently overlooked when quantitative BR analyses are performed. Experimental evidence from clinical trials is handled in empirical manner, which disregards the (pathophysiological or pharmacological) mechanisms associated with the underlying correlations or interdependencies. PKPD models provide an opportunity to quantify such correlations and account for them when performing BR analysis and drawing conclusions about the benefit-risk profile of an intervention. It is also worth mentioning that understanding of the dynamics of disease and its progression is critical to assess the long-term implications of a therapeutic intervention. Such an integrated approach will be illustrated by combining clinical data with available knowledge (e.g. epidemiological data on background rates of expected co-morbidities; or knowledge acquired on a different disease, population or drug of the same class).

More specifically, in the context of chronic iron overload serum ferritin levels are often used as markers of total body iron accumulation. Despite known limitations of instantaneous serum ferritin levels as a predictor of iron organ accumulation, model-based approaches can be developed which incorporate MRI data as well as other measurements (e.g., SQUID or LIC) to better describe tissue specific accumulation (see **paragraph 2.2.2**). However, a challenge remains in that such measurements may not be easily performed or feasible in young patients. Therefore, situations exist in which decision-making will have to be guided by evidence arising from endpoints which do not reflect drug-disease interaction in the target population. An attempt will be made to demonstrate how evidence synthesis by modelling and simulation may provide a more robust basis for extrapolating findings from adults to children and for translating short-term results into long-term predictions.

In **chapter 6** we develop a disease model based on available literature data, in which changes in serum ferritin levels are correlated with RBC transfusion regimen. The approach is developed in a stepwise manner; first we evaluate basal, physiological changes in serum ferritin in healthy individuals by means of a turnover model. Then, the effect of RBC transfusions is added into the model to quantify changes in the production rate of ferritin. Our investigation provides for the first time in a parametric way, evidence of the relationship between blood transfusions and serum ferritin levels. This physiological turnover model forms the basis for a more structured evaluation of chelation therapy in transfusion-dependent iron overload. The approach is subsequently validated in **chapter 7**, where data

from 27 patients affected by transfusion-dependent diseases are used to predict the effects of deferoxamine on ferritin levels.

The scope of the drug-disease model for iron overload is not only to establish the relevance of ferritin levels as a measure of effective chelation therapy. The ultimate goal will be to demonstrate its value as a tool to support decision making in benefit-risk analysis. Of note is the opportunity to explore different scenarios in addition to available clinical evidence. Such scenarios may provide further insight into the role of differences in patient population characteristics and dosing regimens on treatment response as well as enable one to predict potential long-term complications based on short-term effects. Given the multidimensional nature of benefit-risk profile, our approach involves not only the integration and parameterisation of a drug-disease model for efficacy measures, but also for safety endpoints. Therefore, in **chapter 8**, we evaluate the acute and long-term complications of iron chelation therapy using the data obtained from patients undergoing chelation with deferoxamine. Different adverse events are considered, which reflect typical features of adverse drug reactions, including short and long term events, as well as dose-dependent and dose-independent effects. Such a comprehensive analysis is proposed by integrating epidemiological (literature) and pharmacological data. In doing so, we also ensure that interdependencies and correlations between the different endpoints under evaluation are taken into account in a quantitative manner.

As in many other chronic diseases, compliance to the prescribed dose and dosing regimen is an important factor in chelation therapy. We illustrate how patient behaviour regarding compliance to treatment contributes to changes in ferritin levels and consequently affect the overall benefit-risk profile of an intervention. Simulation scenarios are evaluated in which different compliance patterns are used to assess changes in the magnitude and incidence of acute and long-term complications.

2.5 Clinical trial and not-in-trial simulations: accounting for exposure, disease progression and uncertainty in benefit-risk analysis

Throughout this thesis we have hypothesised that model-guided evidence generation and subsequent integration of new clinical data with available knowledge (i.e., evidence synthesis) provides a robust framework for characterising the benefit-risk profile of any intervention. We also highlight the limitations of current practice in that any attempt to establish the benefit-risk profile at the moment of drug approval relies only on the evidence generated (e.g. treatment arms tested throughout the drug development phases). Such an approach presupposes that the available data are representative of the response profile in

target population and suffice to support key decisions about the favourable benefit-risk profile and suitability of the recommended dose and dosing regimens. The underlying assumptions appear to overlook the fact that in children the natural time course of disease occurs in parallel to developmental (physiological) growth and maturation processes. The interaction between these processes may lead to significant changes in the benefit-risk profile over time and such changes are not evident at the time of approval, nor necessarily well captured by long term safety monitoring, as implemented in pharmacovigilance plans. Once more we show that a model-based approach can be used in which virtual scenarios are created taking into account clinical trial design features, as well as real life factors which are known to play a role in clinical practice, such as variable compliance patterns. By performing clinical trial simulations and not-in-trial simulations, intrinsic and extrinsic sources of variation as well as confounding factors can be appropriately evaluated and incorporated into the decision process.

Clearly, most of the points-to-consider described in the previous paragraph are currently overlooked or excluded from quantitative BR analysis, independent of the methodology used. The highlight of this thesis is therefore presented in **chapter 9.**, where we illustrate how new evidence (from typical clinical programmes) can be integrated with existing knowledge in a parametric manner, using drug-disease models in the context of clinical trial protocols or real-life use of the drug. This simulation framework provides a more robust basis for establishing the benefit-risk profile of treatment in children. In fact, clinical trial and not-in-trial simulations offer the opportunity to explore scenarios in which the impact of covariate factors can be assessed without being limited only to the data available.

Moreover, we propose the use of Multi Criteria Decision Analysis (MCDA) as the method of choice for evaluating real and virtual data together. As discussed in the introduction of this thesis, MCDA appears to have the necessary features to characterise and summarise the BR profile of a treatment in a systematic and transparent manner. In **chapter 9**, we perform MCDA to establish the benefit-risk profile of iron chelation therapy with deferoxamine in thalassaemic patients undergoing frequent transfusions. The drug-disease models developed in the previous chapters are used to simulate a range of scenarios; describing typical clinical trials and long term follow up. During the analysis the same relative weight is given to both types of data, i.e., the available data from clinical trials and the predicted profiles inferred from the models. A standard phase III trial (“real data”) is used as a reference scenario and a number of alternative dosing algorithms are proposed and compared (“virtual data”). For the sake of clarity, here we only look at the optimisation of the dosing regimen and how the different options proposed can influence the BR profile. The intent of these scenarios is to illustrate how drug-disease models in conjunction with simulations can better support regulatory and clinical decision making. A range of applications can be considered, in that

the proposed simulation framework could also be used to optimise study design before the implementation of clinical trials. But most importantly, it could form the basis for personalised medicines. Clinical trial and not-in-trial simulations allow us to quantify the impact of relevant covariate factors on treatment outcome, thereby demonstrating the implications of treatment and population stratification.

2.6 Conclusions and perspectives

The results and conclusions drawn from our research are summarised and discussed in **chapter 10**. In this concluding chapter, we revisit the different examples presented in this thesis and we attempt to shed light on the issues currently faced by clinicians, sponsors and regulators involved with the evaluation of the benefit risk profile of a treatment. We make clear that evidence generation has been the paradigm for the development and approval of new drugs. This paradigm is inefficient and should be questioned for a number of reasons. The assumption that arising evidence from clinical trials discriminates drug-specific properties from the underlying progression of disease overlooks shortcomings such as limited accuracy and precision of the estimates for endpoints, which will be subsequently used for BR assessment.

We defend the need for a development and approval paradigm which relies on a framework which supports evidence generation and evidence synthesis as the basis for approval. Clinical events or the absence thereof are not spurious, random features of an intervention. They are greatly determined by the patient population, the context in which the treatment is assessed and by the dose rationale. In addition, we emphasise in this last section, how clinical trial and not-in-trial simulations can be used to complement clinical evidence. We envisage that such a framework will provide a more structured basis for BR analysis, reducing uncertainties about the changes in benefit-risk profile, which are intrinsic to the progression of disease and take place in parallel to maturation developmental growth in children.

We conclude this thesis with a set of answers to longstanding clinical questions regarding the use of iron chelators in chronic iron overload. The approach used to address those questions also highlights opportunities for future research in quantitative pharmacology, especially with regard to the development of multidimensional models and the relevance of Bayesian statistical inference for the implementation of such models. In our final remarks we include suggestions regarding the requirements for the prospective implementation of this framework as a tool for regulatory approval and risk management for paediatric medicines.

References

1. Food and Drug Administration (FDA). Guidance for Industry Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products. 1998;(May):1–23.
2. Food and Drug Administration (FDA). Guidance for Industry: Exposure-Response Relationships - Study Design, Data Analysis and Regulatory Applications. 2003;(April):1–25.
3. Downing NS, Aminawung J a, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA*. 2014;311(4):368–77.
4. Katz R. FDA: evidentiary standards for drug development and approval. *NeuroRx*. 2004;1(July):307–16.
5. European Medicines Agency (EMA). Benefit-risk methodology project - Project description. 2009.
6. FDA. Structured Approach to Benefit-Risk Assessment in Drug Regulatory Decision-Making - Draft PDUFA V Implementation Plan. 2013.
7. Liberti B, McAuslane N, Walker S. Progress on the Development of a Benefit / Risk Framework for Evaluating Medicines. *Regul Focus*. 2010;15:32–7.
8. Cohen J, Neumann P. What’s more dangerous, your aspirin or your car? Thinking rationally about drug risks (and benefits). *Heal Aff*. 2007;26:636–46.
9. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. *Value Health*. 2010 Aug;13(5):657–66.
10. Holden WL. Benefit-risk analysis : a brief review and proposed quantitative approaches. *Drug Saf*. 2003 Jan;26(12):853–62.
11. ICH Topic E11.
http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002926.pdf. 2001.
12. EMA. Paediatric Workshops -
http://www.emea.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000416.jsp&mid=WC0b01ac0580118a19.
13. Manolis E, Pons G. Proposals for model-based paediatric medicinal development within the current European Union regulatory framework. *Br J Clin Pharmacol*. 2009 Oct;68(4):493–501.
14. Sammons HM, Starkey ES. Ethical issues of clinical trials in children. *Paediatr Child Health (Oxford)*. Elsevier Ltd; 2012 Feb;22(2):47–50.

15. EMA. Benefit-risk methodology project Work package 4 report : Benefit-risk tools and processes. 2012.
16. Gibbons R, Higgs DR, Old JM, Olivieri NF, Swee Lay T, Wood WG. The Thalassaemia Syndromes - Fourth Edition. Blackwell Sci. 2001;
17. Thalassaemia International Federation (TIF). Guidelines for the clinical management of thalassaemia. 2008.
18. Galanello R, Origa R. Beta-thalassemia. *Orphanet J Rare Dis.* 2010 Jan;5:11.
19. Andrews N. Disorders of Iron Metabolism. *N Engl J Med.* 1999;341(26):1986–95.
20. Andrews NC, Schmidt PJ. Iron homeostasis. *Annu Rev Physiol.* 2007 Jan;69:69–85.
21. Anderson L. Cardiovascular T2-star (T2*) magnetic resonance for the early diagnosis of myocardial iron overload. *Eur Heart J.* 2001 Dec 1;22(23):2171–9.
22. Iancu T. Ferritin and hemosiderin in pathological tissues. *Electron Microsc rev.* 1992;5:209–29.
23. Cazzola M, Huebers H, Sayers M, MacPhail A, Eng M, Finch C. Transferrin saturation, plasma iron turnover, and ferritin uptake in normal humans. *Blood.* 2011;66:935–9.
24. Cook JD, Barry WE, Hershko C, Fillet G, Finch C. Iron Kinetics with emphasis on iron overload. *Am J Pathol.* 1972;337–44.
25. Halliday JW, Powell LW. Ferritin and cellular iron metabolism. *Ann N Y Acad Sci.* 1988 Jan;526:101–12.
26. Knovich MA, Storey JA, Coffman LG, Torti S V. Ferritin for the Clinician. *Blood Rev.* 2010;23(3):95–104.
27. Borgna-Pignatti C, Cappellini MD, De Stefano P, Del Vecchio GC, Forni GL, Gamberini MR, et al. Survival and complications in thalassemia. *Ann N Y Acad Sci.* 2005 Jan;1054:40–7.
28. Borgna-Pignatti C, Rugolotto S, De Stefano P, Zhao H, Cappellini MD, Del Vecchio G, et al. Survival and complications in patients with thalassemia major treated with transfusion and deferoxamine. *Haematologica.* 2004;89(10):1187–93.
29. Borgna-Pignatti C, Cappellini MD, Stefano P De, Carlo G, Vecchio D, Forni GL, et al. Cardiac morbidity and mortality in deferoxamine- or deferiprone-treated patients with thalassemia major. 2009;107(9):3733–7.
30. Cunningham MJ, Macklin E a, Neufeld EJ, Cohen AR. Complications of beta-thalassemia major in North America. *Blood.* 2004 Jul 1;104(1):34–9.

CHAPTER 2

31. Hershko C. Pathogenesis and management of iron toxicity in thalassemia. *Ann N Y Acad Sci.* 2010 Aug;1202:1–9.
32. Cappellini MD, Pattoneri P. Oral iron chelators. *Annu Rev Med.* 2009 Jan;60:25–38.
33. Kwiatkowski JL. Oral iron chelators. *Pediatr Clin North Am.* 2008 Apr;55(2):461–82.
34. Musallam KM, Taher AT. Iron chelation therapy for transfusional iron overload: a swift evolution. *Hemoglobin.* 2011 Jan;35(5-6):565–73.
35. Shander A, Sweeney J. Overview of Current Treatment Regimens in Iron Chelation Therapy. *US Hematol.* 2009;2(1):56–9.
36. Theil EC. Mining ferritin iron: 2 pathways. *Blood.* 2009 Nov 12;114(20):4325–6.
37. Brittenham GM, Cohen a R, McLaren CE, Martin MB, Griffith PM, Nienhuis a W, et al. Hepatic iron stores and plasma ferritin concentration in patients with sickle cell anemia and thalassemia major. *Am J Hematol.* 1993 Jan;42(1):81–5.
38. Lipschitz D, Cook J, Finch C. A clinical evaluation of serum ferritin as an index of iron stores. *N Engl J Med.* 1974;290(22):1213–6.
39. Angelucci E, Brittenham G, McLaren C, Ripalti M, Baronciani D, Giardini C, et al. Hepatic iron concentration and total body iron stores in thalassemia major. *N Engl J Med.* 2000;343(5):327–31.
40. Villeneuve JP, Bilodeau M, Lepage R, Côté J, Lefebvre M. Variability in hepatic iron concentration measurement from needle-biopsy specimens. *J Hepatol.* 1996 Aug;25(2):172–7.
41. Fischer R, Longo F, Nielsen P, Engelhardt R, Hider RC, Piga A. Monitoring long-term efficacy of iron chelation therapy by deferiprone and desferrioxamine in patients with beta-thalassaemia major: application of SQUID biomagnetic liver susceptometry. *Br J Haematol.* 2003 Jun;121(6):938–48.
42. Kirk P, Roughton M, Porter JB, Walker JM, Tanner M a, Patel J, et al. Cardiac T2* magnetic resonance for prediction of cardiac complications in thalassemia major. *Circulation.* 2009 Nov 17;120(20):1961–8.
43. St Pierre TG, Clark PR, Chua-anusorn W, Fleming AJ, Jeffrey GP, Olynyk JK, et al. Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance. *Blood.* 2005 Jan 15;105(2):855–61.

SECTION II

Optimising evidence generation in paediatric trials

CHAPTER 3

Population pharmacokinetics of deferiprone in healthy subjects

Francesco Bellanti, Meindert Danhof, and Oscar Della Pasqua

Br J Clin Pharmacol - 2014

Summary

Aims: To characterise the pharmacokinetics of deferiprone in healthy subjects using a model-based approach and assess the effect of demographic and physiological factors on drug exposure.

Methods: Data from 55 adult healthy subjects receiving deferiprone (solution 100 mg/ml) were used for model building purposes. A population pharmacokinetic analysis was performed using NONMEM VII. The contribution of gender, age, weight, and creatinine clearance (CLCR) on drug disposition was evaluated according to standard forward inclusion, backward deletion procedures. Model selection criteria were based on graphical and statistical summaries.

*Results: A one-compartment model with first order oral absorption was found to best describe the pharmacokinetics of deferiprone. Simulated AUC and C_{max} (respectively mean of 45.80 mg*h/L and 17.67 mg/L after 25 mg/kg single dose and 137.40 mg*h/L and 26.50 mg/L after 75 mg/kg b.i.d.) were comparable with literature references. Gender differences in the apparent volume of distribution (20%) have been identified, which may contribute to an increase in peak concentrations in females. Furthermore, simulation scenarios reveal that dose adjustment is required for patients with reduced CLCR. Doses of 60, 40 and 25 mg/kg for patients showing mild, moderate and severe renal impairment are proposed based on CLCR values of 60-89, 30-59 and 15-29 ml/min, respectively.*

Conclusions: Our analysis has enabled the assessment of the impact of gender and CLCR on the pharmacokinetics of deferiprone. Moreover, it provides the basis for dosing recommendations in renal impairment. The implication of these covariates on systemic exposure is currently not available in the prescribing information of deferiprone.

3.1 Introduction

Patients with β -thalassaemia and other transfusion-dependent diseases develop iron overload from chronic blood transfusions and require regular continuous iron chelation to prevent potentially fatal iron-related complications (1–5). Deferiprone (DFP) is the most extensively studied oral iron chelator to date. DFP is a hydroxypyridone derivative, which was authorised in Europe in 1999 for the treatment of iron overload in patients with β -thalassaemia major when deferoxamine (DFO) is contraindicated or inadequate.

Despite the wide clinical experience with DFP, its pharmacokinetics has not been fully characterised in patients. In addition, there are still limited experimental data available on DFP in children and no data in children under 6 years of age, where the drug is still used off-label. Thus far, it has been established that when administered orally, DFP is rapidly and completely absorbed. Plasma levels show peak concentrations (C_{max}) within 1 hour of administration. Food reduces its absorption rate without affecting the overall exposure to the drug. In patients with β -thalassaemia, the administration of deferiprone at doses of 75 mg/kg/day as a twice-daily regimen yields C_{max} of 34.6 mg/L and area under the plasma concentration-time curve (AUC) of 137.5 mg/L • h (6,7). On the other hand, peak serum concentrations were 17.53 mg/L and 11.82 mg/L in fasting and fed state, respectively after a dose of 25 mg/kg (8). DFP is for the most part inactivated by glucuronidation (>85%) and more than 90% of the drug is removed from plasma within 6 hours of ingestion, with an elimination half-life of 1 to 2.5 hours in patients affected by β -thalassaemia (5,6,9–14). DFP forms a 3:1 complex with iron, which is removed mainly through the kidneys in a similar manner as for the free parent drug. The area under the curve (AUC) of free deferiprone in patients shows high inter-individual variability, which may be related to the variation in the therapeutic response (5,10–12).

The impact of demographic and other physiological factors on the exposure of DFP has not been assessed thus far. In addition, the consequences of such factors for the dosing regimen have not been described in the published literature or on the SmPC (Summary of Product Characteristics) of the drug. Moreover, no information on dose adjustment requirements is provided for patients with hepatic or renal impairment. Given the fast renal elimination of the glucuronide metabolite, renal function is expected to play a major role in affecting the overall exposure to the parent drug.

The aim of this analysis is to characterise the DFP pharmacokinetics in healthy subjects using a model-based approach and assess the effect of demographic and physiological factors on drug exposure. Furthermore, it is our endeavour to show the clinical relevance of simulation scenarios to evaluate the impact of renal impairment on drug disposition and consequently,

for the optimisation of the dosing regimen in special populations. Moreover, we anticipate that the availability of population pharmacokinetic model for deferiprone will facilitate the evaluation of extrapolation of pharmacokinetic data from adults to children. More specifically, it will provide the basis for pharmacokinetic bridging of the dosing regimen for the paediatric population.

3.2 Methods

Data

The pharmacokinetics of deferiprone was evaluated using data collected from two clinical studies: LA20-BA and LA21-BE (15,16), in which healthy subjects received a single dose of 1500 mg of DFP as a 100 mg/ml solution. The studies have been conducted in full conformance with the principles of the Declaration of Helsinki and with the local laws and regulations concerning clinical trials. The protocol and the informed consent documents for each study have been formally approved by the relevant research ethics committee of each clinical site. The data was supplied by ApoPharma Inc, Canada and shared within the DEEP consortium (www.deep.cvbf.net). The DEEP consortium addresses an EU call with the objective of increasing the knowledge of deferiprone chelation therapy in the paediatric population.

Both study protocols were approved by Ethics Committee and all experimental procedures performed according to good clinical practice guidelines. In brief, 55 adult healthy subjects (39 males and 16 females) who had received the active medication were included in the analysis. Blood samples for the evaluation of deferiprone concentrations were taken before and at the following sampling times after dosing: 0.167, 0.333, 0.5, 0.75, 1, 1.333, 1.5, 1.667, 2, 2.5, 3, 4, 5, 6, 8, 10, and 14 hours. On average, 15 samples were collected per subject. Median (range) age (years) and body weight (kg) of the adult population were 39 (19-55) and 72 (52-92) respectively.

Bioanalysis

Deferiprone plasma concentrations were analysed by a validated method previously developed by ApoPharma (Toronto, Canada) using high performance liquid chromatography with UV detection (HPLC-UV). Extraction of deferiprone from supernatant was performed after precipitation of plasma proteins by trichloroacetic acid (TCA - 15%) and centrifugation at 10,000 g for 20 minutes at 4 °C. The analytical column used for the analysis was a Hamilton PRP-1 and separation of the chromatogram of interest was achieved using an isocratic mobile phase (pH 7.0). The UV detector was set at 280 nm. In a recent review of the method, calibration, accuracy and precision estimates have been revisited by our group. The analytical range was between 3.13 and 800 µM (equivalent to 0.43 to 111 µg/ml); and an R² value greater than 0.98 was required to accept the standard curve. The lower limit of

quantification (LLOQ) was 1 μM (equivalent to 0.14 $\mu\text{g}/\text{ml}$). Inter- and Intra-day accuracy and precision were found to be always below 10%, subsequently matching the GLP validation criteria (17).

Pharmacokinetic Modelling

Nonlinear mixed effects modelling was performed in NONMEM version 7.2 (Icon Development Solutions, USA). Model building criteria included: (i) successful minimisation, (ii) standard error of estimates, (iii) number of significant digits, (iv) termination of the covariance step, (v) correlation between model parameters and (vi) acceptable gradients at the last iteration.

Fixed and random effects were introduced into the model in a stepwise manner. Inter-individual variability in pharmacokinetic parameters was assumed to be log-normally distributed. A parameter value of an individual i (post hoc value) is therefore given by the following equation:

$$\theta_i = \theta_{TV} * e^{\eta_i}$$

in which θ_{TV} is the typical value of the parameter in the population and η_i is assumed to be random variable with zero mean and variance ω^2 . Residual variability, which comprises measurement and model error, was described with a proportional error model. This means for the j th observed concentration of the i th individual, the relation Y_{ij} :

$$Y_{ij} = F_{ij} + \varepsilon_{ij} * W$$

where F_{ij} is the predicted concentration and ε_{ij} the random variable with mean zero and variance σ^2 . W is a proportional weighing factor for ε .

Goodness of fit was assessed by graphical methods, including population and individual predicted vs. observed concentrations, conditional weighted residual vs. observed concentrations and time, correlation matrix for fixed vs. random effects, correlation matrix between parameters and covariates and normalised predictive distribution error (NPDE) (18,19). Comparison of hierarchical models was based on the likelihood ratio test. A superior model was also expected to reduce inter-subject variability terms and/or residual error terms.

Covariate analysis

Continuous and categorical covariates were tested during the analysis. The relationship between individual PK parameters (post-hoc or conditional estimates) and covariates was explored by graphical methods (plot of each covariate vs. each individual parameter). Relevant demographic covariates (body weight, age, gender, creatinine clearance) were entered one by one into the population model (univariate analysis). After all significant covariates had been entered into the model (forward selection), each covariate was removed (backward elimination), one at a time. The model was run again and the objective

function recorded. The likelihood ratio test was used to assess whether the difference in the objective function between the base model and the full (more complex) model was significant. The difference in -2Log likelihood (DOBJF) between the base and the full model is approximately χ^2 distributed, with degrees of freedom equal to the difference in number of parameters between the two hierarchical models. Because of the exploratory nature of this investigation, for univariate analyses, additional parameters leading to a decrease in the objective function of 3.84 was considered significant ($p < 0.05$). During the final steps of the model building, only the covariates which resulted in a difference of objective function of at least 7.88 ($p < 0.005$) were kept in the final model.

Model validation

The validation of the final pharmacokinetic model was based on graphical and statistical methods, including visual predictive checks (15). Given the importance of the validation procedures for the subsequent use of a model for simulation purposes, in this study we have included a wide range of diagnostic methods to assess the accuracy of the parameter estimates and the predictive performance of the model (16). Bootstrap was used to identify bias, stability and accuracy of the parameter estimates (standard errors and confidence intervals). The bootstrap procedures were performed in PsN v3.5.3 (University of Uppsala, Sweden) (20), which automatically generates a series of new data sets by sampling individuals with replacement from the original data pool, fitting the model to each new data set. Subsequently, parameter estimates were used to simulate plasma concentrations in subjects with similar demographic characteristics, dosing regimens and sampling scheme as in the original clinical studies. Mirror plots were also generated to evaluate the variance-covariance structure of the parameters in the model, which is reflected by the degree of similarity between the original fit and the pattern obtained from the fitting of the simulated data sets using the final pharmacokinetic model.

In addition to the graphical analysis, posterior predictive check was performed using AUC (area under the plasma concentration vs. time curve) and C_{max} (peak plasma concentration) as a measure of model performance. AUC and C_{max} values were calculated non-compartmentally by trapezoidal method from simulations of 1000 data sets with the same demographic characteristics, dosing regimens and sampling scheme as in the original clinical studies.

The distribution of model-predicted AUC and C_{max} values were presented for geometric mean, lower and upper boundaries of the 95% confidence intervals and compared to the findings from non-compartmental analysis in the two clinical studies. Model performance was demonstrated by the location of the original estimates across the predicted distribution (histograms).

Simulation scenarios

Simulations were performed using the final model to assess whether predicted secondary PK parameters, such as AUC and C_{MAX} would be in line with literature references (7,14,21). 30 simulated patients (15 males and 15 females) with a mean body weight of 55 kg (sd 8.4) received DFP under the following dosing recommendations: 25 and 75 mg/kg/day.

Furthermore, additional simulation scenarios were evaluated to assess the implications of renal impairment for the pharmacokinetics of deferiprone in a group of patients with similar demographic characteristics, as described above. Taking into account the correlation between the reduction in creatinine clearance and the severity of renal impairment, three scenarios were considered, including 80, 50 and 25% of the normal clearance values. They were meant to reflect the changes in renal function in mild, moderate and severe impairment, respectively. Simulated patients received 75 mg/kg/day DFP and their exposure was compared to healthy subjects (reference population). Dosing regimens were adjusted to ensure that deferiprone exposure similar to the levels observed in the reference population is achieved and maintained irrespective of the degree of renal impairment.

3.3 Results

Population Pharmacokinetic Modelling

The pharmacokinetics of DFP was best described by a one-compartment model with first-order absorption, lag-time to central compartment, and first-order elimination. Inter-individual variability (IIV) could be estimated for apparent clearance (CL/F), apparent volume of distribution (V/F), and absorption rate constant (K_a). Residual variability was characterised by a proportional error model with a weighting factor.

During covariate model selection, the effect of age, gender and body weight was tested on relevant pharmacokinetic parameters. Initially when tested separately, significant effects of gender on V/F and body weight on CL/F and V/F were identified and described according to a linear model. However, despite statistical significance and improvement in the goodness-of-fit, the inclusion of body weight on either CL/F or V/F also led to an important reduction in model stability during bootstrapping procedures, which is likely caused by the limited range of the covariate values in the study population. Therefore, only gender on V/F was retained in the final model. This resulted in a better description of the data, subsequently increasing the model performance. An overview of the parameter estimates is presented in Table 1.

Table 1. Population pharmacokinetic parameters of deferiprone and bootstrap results.

Parameters	Final model	Bootstrap = 500 runs	
	Estimate	Median	CV (%)
CL/F (L/h)	30.8	30.9	3.12
V/F males (L)	78.4	78.53	2.39
Ka (h ⁻¹)	8.2	8.73	29.2
Lagtime (h ⁻¹)	0.146	0.145	3.93
Error: weighting factor	2.4	2.41	15.26
V/F females (L)	65.3	65.3	3.88
Eta CL/F (%)	0.057 (23.87 %)	0.0557 (23.6 %)	17.59
Eta V/F (%)	0.0278 (16.67 %)	0.0267 (16.34 %)	20.22
Block CL-V	0.0345	0.0335	20
Eta Ka (%)	0.991 (99.54 %)	1.00 (100 %)	23.8
Sigma (%)	0.00566 (7.52 %)	0.00568 (7.53 %)	25.88

Internal model validation diagnostics were satisfactory. Individual predicted profiles and goodness-of-fit plots reveal that the model provides an adequate and non-biased description of the data, as shown in Figure 1 and 1S (see supplemental material in appendix). In addition, despite a small deviation at the tails of the distribution, NPDE summaries (Figure 2S, see supplemental material in appendix) show that the discrepancy between predicted and observed values can be assumed to be normally distributed.

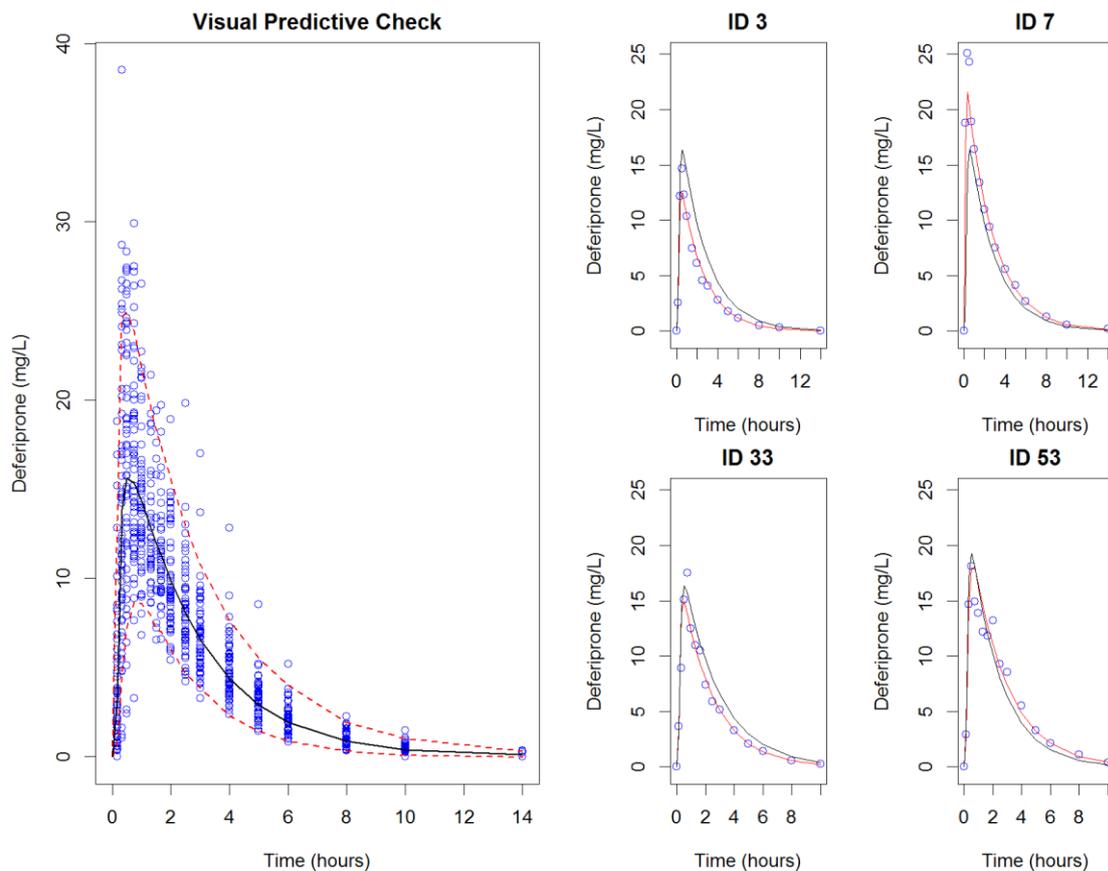


Figure 1. Visual Predictive Check and a random selection of individual plots. VPC on the left panel: observed data are plotted using blue circles; the black solid line represents the median of the simulated data; the red solid lines represent the 5th and 95th percentiles of the simulated data. Individual plots of 4 randomly selected patients: observed data are plotted using blue circles; the black solid line represents the population prediction (Pred) and the red solid line represents the individual predictions (IPred).

The predictive performance of the model in subsequent simulations was deemed critical to achieve the objective of our analysis. To this purpose, mirror plots were therefore used to assess whether the variance and covariance structures have been well characterised. Lastly, the median parameter estimates from the bootstrap analysis were found to be in close agreement with the results observed during the original fitting. Results from the bootstrap analysis are presented in Table 1. Overall these diagnostic techniques confirm that the final model is suitable for the purposes of data simulation.

Simulation scenarios

First an attempt was made to perform external validation of the model by deriving secondary parameters (AUC and C_{max}) and comparing model-predicted estimates literature references (7,14,21). As shown in Figure 2, reference values lie within the distribution of

simulated AUC and Cmax, for which the mean and 90% CI were 45.80 (44.42-47.17) mg*h/L and 17.67 (17.13-18.20) mg/L, respectively after administration of a single oral dose of 25 mg/kg deferiprone and 137.40 (133.27-141.52) mg*h/L and 26.50 (25.70-27.29) mg/L, respectively after administration of 75 mg/kg/day dose as a twice-daily regimen. Despite the gender differences in the volume of distribution, no significant differences observed when comparing Cmax values. This may be explained by the limited number of females in our analysis as well as by the differences in deferiprone formulation used in past protocols.

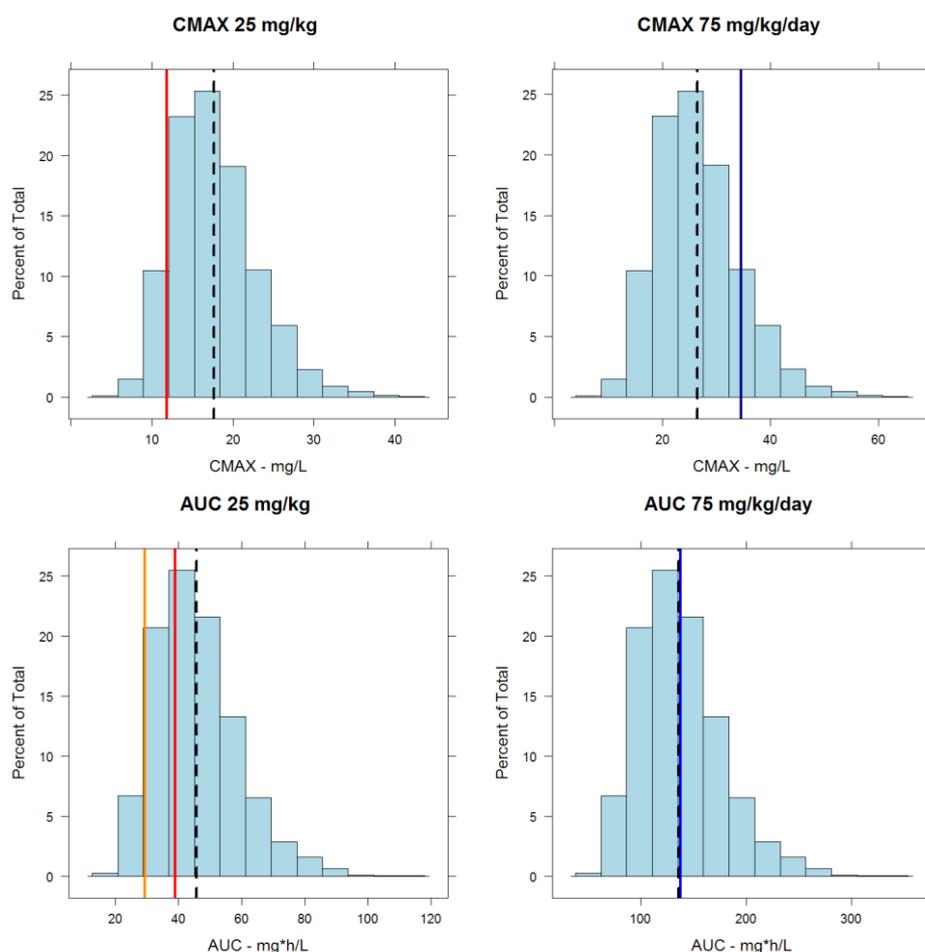


Figure 2. Comparison of secondary PK parameters (Cmax and AUC) with literature references. Predicted DFP exposure expressed as CMAX and AUC for adult patients receiving 25 mg/kg as a single dose and 75 mg/kg/day b.i.d. The dashed black lines depict the mean simulated values, whereas the solid coloured lines depict published results (7, 14, 17). Percent of total indicates the percentage of cases for each beam of 100 simulations with 55 patients in each simulated trial.

As the population available for the analysis was limited to healthy volunteers, the impact of another important covariate could not be estimated during the fitting procedures, namely, the role of glomerular filtration as determined by changes in creatinine clearance. Therefore

a simulation-based approach was used to quantify the implications of renal impairment for the disposition of deferiprone. Systemic exposure expressed as AUC was simulated for three scenarios representing mild, moderate and severe impairment and compared to the estimates obtained for healthy subjects. It is evident from Figure 3 that over-exposure occurs when comparing the three sub-populations receiving 75 mg/kg/day DFP with the reference data, particularly in the case of moderate and severe impairment. Given the magnitude of the increase in systemic exposure, dose adjustment should be recommended for patients with renal impairment.

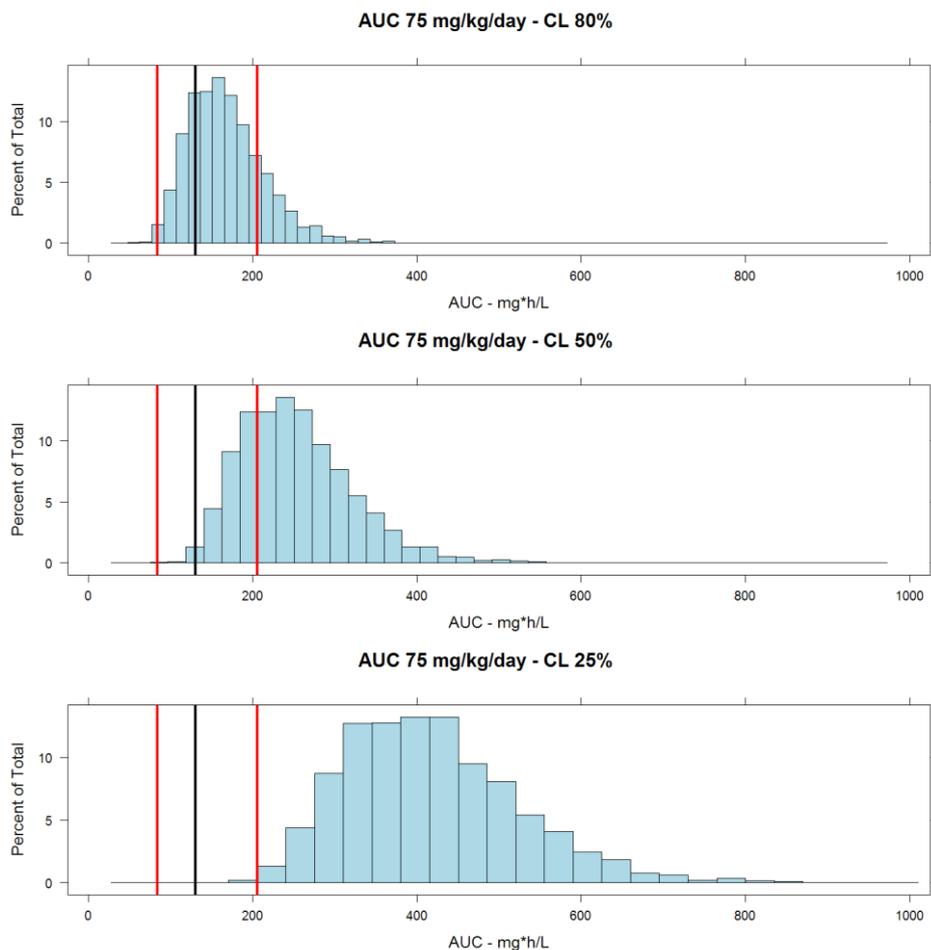


Figure 3. AUC distributions: 80, 50 and 25% of total clearance (DFP 75 mg/kg/day). Predicted DFP exposure expressed as AUC for adult patients receiving 75 mg/kg/day and presenting 80%, 50% and 25% of the total clearance respectively. The black line represents the median of the reference population which presents normal renal function, whereas the red lines represent 5th and 95th percentiles of the same reference population. Percent of total indicates the percentage of cases for each beam of 100 simulations with 55 patients in each simulated trial.

As shown in Figure 4, dose adjustments can be considered that allow for deferiprone exposure to be maintained at the desired levels for all three scenarios. In addition, Figure 5 depicts the consequence of reduced clearance for the systemic exposure of deferiprone assuming first-order pharmacokinetics in this population. Doses of 60, 40 and 25 mg/kg for patients showing mild, moderate and severe renal impairment are proposed based on creatinine clearance values of 60-89, 30-59 and 15-29 ml/min, respectively. An overview of these recommendations is summarised in Table 2.

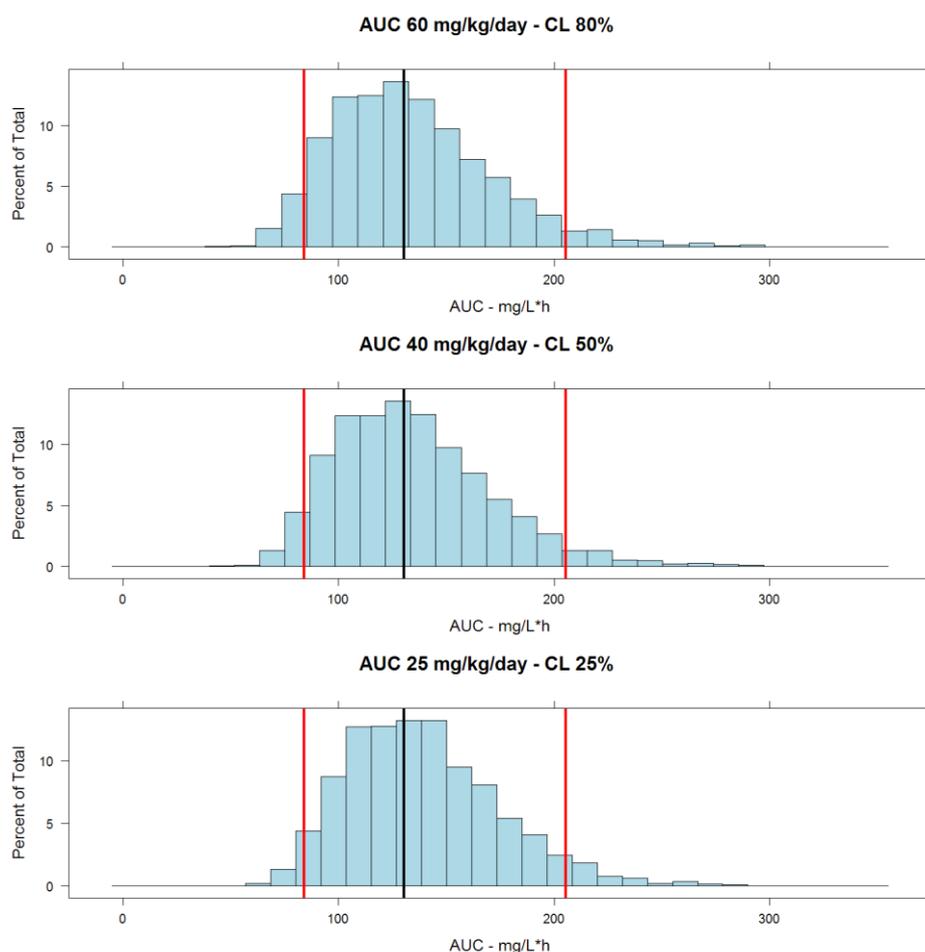


Figure 4. AUC distributions: 80, 50 and 25% of total clearance (new dosing recommendations). Predicted DFP exposure expressed as AUC for adult patients receiving the adjusted dosing recommendation based on the severity of renal impairment. The three populations present 80%, 50% and 25% of the total clearance respectively. The black line represents the median of the reference population which presents normal renal function, whereas the red lines represent 5th and 95th percentiles of the same reference population. Percent of total indicates the percentage of cases for each beam of 100 simulations with 55 patients in each simulated trial.

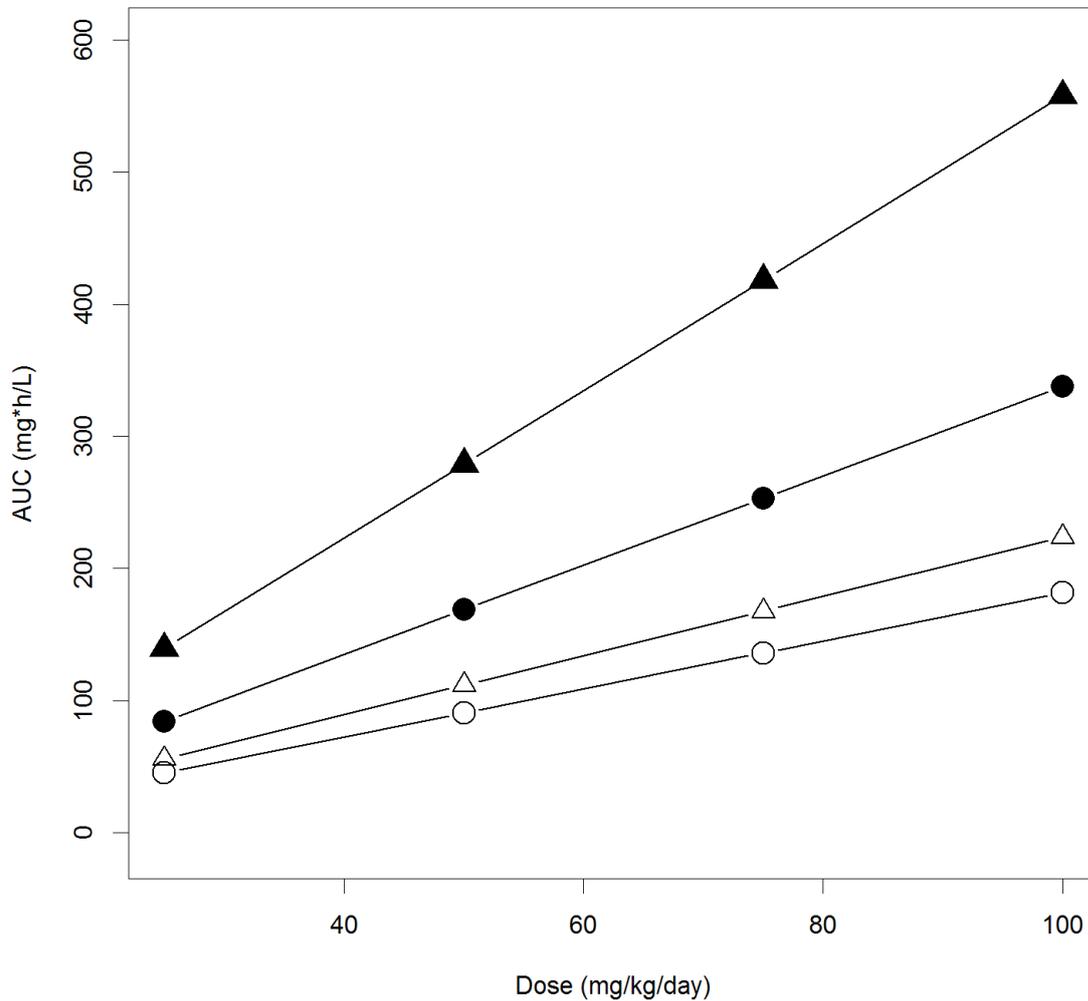
AUC-DOSE relationship in the presence of renal impairment

Figure 5. AUC – Dose relationship: 80, 50 and 25% of total clearance. AUC – Dose relationship in the presence of renal impairment. The open circles represent the reference population with normal renal function. The open triangles, filled circles and filled triangles represent mild (80%), moderate (50%) and severe (25%) renal impairment respectively.

Table 2. New dosing recommendations for renal impairment.

Scenarios	Degree of impairment	CrCL (ml/min)	Standard dosing recommendations	Proposed dosing recommendations
80% of total Clearance	Mild	60-89	75 mg/kg/day	60 mg/kg/day
50% of total Clearance	Moderate	30-59	75 mg/kg/day	40 mg/kg/day
25% of total Clearance	Severe	15-29	75 mg/kg/day	25 mg/kg/day

3.4 Discussion and Conclusion

As generally known, inter-individual variability in PK can significantly affect the outcome of a given therapeutic intervention. Therefore, full optimisation of the therapeutic regimen cannot be achieved without taking variability into account. The use of model-based approaches for the evaluation of the dose rationale and personalisation of dosing regimens for subgroups of patients and special populations has become an invaluable tool as it allows characterisation and quantification of the contribution of different sources of variability to the overall pharmacokinetic properties. This has an even larger impact when considering special populations and rare diseases, as is the case of transfusion dependent diseases and other pathologies associated with renal and hepatic impairment. Despite the continuous emphasis on the need for evidence-based clinical and regulatory decisions, modelling and simulation is becoming an essential component of evidence synthesis, which ultimately underpins decisions and recommendations (22–24).

Deferiprone Pharmacokinetics

With this analysis we show how population pharmacokinetics can be used to explore the implications of different sources of variability on the exposure of the oral iron chelator deferiprone. The estimates of the main parameters of interest (table 1) were in line with previously published results (6,7,10–13,21,25–27). As shown in figure 2, similar agreement was also observed for the secondary PK parameters (AUC and C_{max}). By contrast, no gender differences have been identified in previous studies. In this respect, our analysis illustrates the importance of parametric methods for accurate evaluation of covariate effects. We have quantified gender differences in the apparent volume of distribution, where V/F was estimated to be 78.4 and 65.3 L in males and females, respectively (i.e., a 20% difference between the two groups). Assuming that overall exposure (AUC) rather than C_{max} is the primary determinant of response, these differences are likely to have minor clinical implications.

Dosing recommendations in patients with renal impairment

Considering the lack of details in the label of DFP regarding the dose rationale for special populations, it was our interest to provide insights on dosing recommendations for patients with renal impairment, which occurs as co-morbidity in thalassaemia. Given that, independently of the metabolic rate, 90% of the total drug (free, metabolised and iron-complex) is excreted in the urine within 5 to 6 hours of ingestion, we have assumed that renal impairment would be clinically more relevant, as compared to hepatic impairment. We have selected a discrete number of scenarios to describe different levels of impairment (mild, moderate and severe). As could be anticipated for any drug with primary renal elimination (28,29), use of the standard recommended dose of 75 mg/kg/day leads to overexposure to deferiprone; especially when clearance is reduced beyond 50% of the normal range. Taking into account the deferiprone levels associated with effective response, dosing regimens are proposed for the three sub-populations allowing exposure to remain comparable to values observed in patients with normal renal function.

A look into the future: rare diseases and special populations

As discussed above, model-based approaches can be critical for therapeutic decisions when limited evidence is available. This is certainly the case for transfusion dependent diseases, especially when considering young paediatric patients, for whom limited data or no data exist and the use of DFP is still off label.

Our analysis represents the first attempt to synthesise current knowledge on the pharmacokinetics of deferiprone and subsequently optimise the dosing regimen in special populations. In addition to renal impairment, we envisage the use of this model for the optimisation of clinical trial design in children. It is worth mentioning that optimisation of protocol design may enable the use of smaller cohorts as well as a considerable reduction in the burden associated with sampling procedures thanks to the use of sparse sampling techniques.

Limitations and Assumptions

Given that the model has been developed on data collected in healthy subjects, questions arise about the relevance of the parameter estimates for the target patient population. No differences have been found in previous analyses between healthy individuals and patients. In the work carried out by Stobie *et al.* (21), who compared the pharmacokinetics of DFP in healthy individuals with patients affected by β -thalassaemia, only a slight difference in the apparent volume of distribution was observed, but the results were found not to be statistically significant (6,21). Most importantly, the authors did not find any differences in the drug clearance between the two groups. Moreover, AUC and C_{max} values simulated by our model (figure 2) were comparable with published data obtained in patients treated with

DFP. We have to acknowledge that a lower mean C_{max} is observed when comparing simulated data and reference data at 75 mg/kg/day b.i.d. This could be the consequence of a difference in the V_d observed in patients and/or differences in the formulation. Having said that, we anticipate that such a change should have limited clinical implications for the following reasons: overall exposure is the determinant of the response and AUC values were comparable between the two groups; an increase in C_{max} is not expected to have consequences from a safety perspective, as discussed also for gender differences; and additionally the recommended dosing regimen is given as a three times daily administration which further reduces the impact of C_{max} changes. We believe therefore that eventual differences in haemodynamics in patient affected by transfusion dependent diseases will not be relevant for the overall disposition properties of deferiprone.

Conclusion

In conclusion, our analysis has allowed the identification of the effect of gender on the volume of distribution of DFP and enabled the evaluation of the dosing requirements for patients with renal impairment. The changes in dose regimen proposed for this special population should be considered when prescribing DFP to this population.

References

1. Clegg JB, Higgs DR, Gibbons R, Old JM, Olivery FN, Swee Lay T, Weatherall DJ, Wood WG. The pathophysiology of the thalassaemias. *In: The Thalassaemia Syndromes*. 4th edition. Gibbons R, Higgs DR, Old JM, Olivieri NF, Swee Lay T, Wood WG. (eds) Blackwell Sci. Oxford, 2001; 192-236
2. Cunningham MJ, Macklin EA, Neufeld EJ, Cohen AR. Thalassaemia Clinical Research Network. Complications of beta-thalassaemia major in North America. *Blood*. 2004; 104(1):34-9.
3. Borgna-Pignatti C, Cappellini MD, De Stefano P, Del Vecchio GC, Forni GL, Gamberini MR, Ghilardi R, Origa R, Piga A, Romeo MA, Zhao H, Cnaan A. Survival and complications in thalassaemia. *Ann NY Acad Sci*. 2005; 1054:40-7
4. Borgna-Pignatti C, Cappellini MD, De Stefano P, Del Vecchio GC, Forni GL, Gamberini MR, Ghilardi R, Piga A, Romeo MA, Zhao H, Cnaan A. Cardiac morbidity and mortality in deferoxamine- or deferiprone-treated patients with thalassaemia major. *Blood*. 2006; 107(9):3733-7.
5. Galanello R, Campus S. Deferiprone chelation therapy for thalassaemia major. *Acta Haematol*. 2009; 122(2-3):155-64.
6. Barman Balfour JA, Foster RH. Deferiprone: a review of its clinical potential in iron overload in beta-thalassaemia major and other transfusion-dependent diseases. *Drugs*. 1999; 58(3):553-78.
7. Fassos FF, Klein J, Fernandes D, Matsui D, Olivieri NF, Koren G. The pharmacokinetics and pharmacodynamics of the oral iron chelator deferiprone (L1) in relation to hemoglobin levels. *Int J Clin Pharmacol Ther*. 1996; 34(7):288-92.
8. European Medicines Agency (EMA). Deferiprone Summary of Product Characteristics. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000236/WC500022050.pdf (last accessed on 20/01/2014)
9. Benoit-Biancamano MO, Connelly J, Villeneuve L, Caron P, Guillemette C. Deferiprone Glucuronidation by Human Tissues and Recombinant UDP Glucuronosyltransferase 1A6: An in Vitro Investigation of Genetic and Splice Variants. *Drug Metab Dispos*. 2009; 37(2):322-9.
10. Diav-citrin O, Koren G. Oral iron chelation with deferiprone. *New Front Pediat Drug Ther*. 1997; 44(1):235-47.
11. Hoffbrand AV, Cohen A, Hershko C. Role of deferiprone in chelation therapy for transfusional iron overload. *Blood*. 2003; 102(1):17-24.
12. Hoffbrand AV. Deferiprone therapy for transfusional iron overload. *Best Pract. Res Clin Haematol*. 2005; 18(2):299-317.

13. Limenta LMG, Jirasomprasert T, Tankaniltert J, Svasti S, Wilairat P, Chantharaksri U, Fucharoen S, Morales NP. UGT1A6 genotype-related pharmacokinetics of deferiprone (L1) in healthy volunteers. *Br J Clin Pharmacol*. 2008; 65(6):908–16.
14. Limenta LMG, Jirasomprasert T, Jittangprasert P, Wilairat P, Yamanont P, Chantharaksri U, Fucharoen S, Morales NP. Pharmacokinetics of Deferiprone in Patients with Beta-thalassemia. *Clin Pharmacokinet*. 2011; 50(1):41–50.
15. Apotex Research Inc. Clinical study report LA20-BA: An open label, single-dose, three-way crossover bioavailability study of deferiprone tablets (Ferriprox) and deferiprone solution under fasting and fed conditions. 2005.
16. ApoPharma Inc. Clinical study report LA21-BE. Randomized, open label, comparative, twoway crossover bioavailability study of deferiprone oral solution and Ferriprox (deferiprone) tablets under fasting conditions. 2005.
17. Bellanti F, Della Pasqua O. Internal report: Analytical procedure for the determination of deferiprone in human plasma by HPLC. 2011.
18. Hooker AC, Staats CE, Karlsson MO. Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method. *Pharm Res*. 2007; 24(12):2187–97.
19. Comets E, Brendel K, Mentré F. Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. *Comput Methods Program Biomed*. 2008; 90(2):154–66.
20. Lindbom L, Ribbing J, Jonsson EN. Perl-speaks-NONMEM (PsN)--a Perl module for NONMEM related programming. *Comput Method Program Biomed*. 2004; 75(2):85–94.
21. Stobie S, Tyberg J, Matsui D, Fernandes D, Klein J, Olivieri N, Bentur Y, Koren G. Comparison of the pharmacokinetics of 1,2-dimethyl-3-hydroxypyrid-4-one (L1) in healthy volunteers, with and without co-administration of ferrous sulfate, to thalassemia patients. *Int J Clin Pharmacol Ther Toxicol*. 1993; 31(12):602-5.
22. Bellanti F, Della Pasqua O. Modelling and simulation as research tools in paediatric drug development. *Eur J Clin Pharmacol*. 2011; 67 Suppl 1:75–86.
23. Manolis E, Osman TE, Herold R, Koenig F, Tomasi P, Vamvakas S, Raymond AS. Role of modeling and simulation in pediatric investigation plans. *Paed Anaesth*. 2011; 21(3):214–21.
24. 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.
25. Al-Refaie FN, Sheppard LN, Nortey P, Wonke B, Hoffbrand AV. Pharmacokinetics of the oral iron chelator deferiprone (L1) in patients with iron overload. *Br J Haematol*. 1995; 89(2):403–8.

26. Matsui D, Klein J, Hermann C, Grunau V, McClelland R, Chung D, St-Louis P, Olivieri N, Koren G. Relationship between the pharmacokinetics and iron excretion pharmacodynamics of the new oral iron chelator 1,2-dimethyl-3-hydroxypyrid-4-one in patients with thalassemia. *Clin Pharmacol Ther.* 1991; 50(3):294-8.
27. Rodrat S, Yamanont P, Tankanilt J, Chantrarak Sri U, Fucharoen S, Morales NP. Comparison of pharmacokinetics and urinary iron excretion of two single doses of deferiprone in β -thalassemia/hemoglobin E patients. *Pharmacology.* 2012 Jan;90(1-2):88–94.
28. Renal Impairment Working Group. U.S. Department of Health and Human Services. Food and Drug Administration. Guidance for Industry Pharmacokinetics in Patients with Impaired Renal Function — Study Design, Data Analysis, and Impact on Dosing and Labeling. 1998. Available from: <http://www.fda.gov/downloads/Drugs/.../Guidances/UCM204959.pdf> (last accessed on 20/01/2014).
29. Efficacy Working Party. Committee for medicinal products for human use (CHMP). European Medicines Agency. Note for guidance on the evaluation of the pharmacokinetics of medicinal products in patients with impaired renal function. 2004. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003123.pdf (last accessed on 20/01/2014).

Appendix

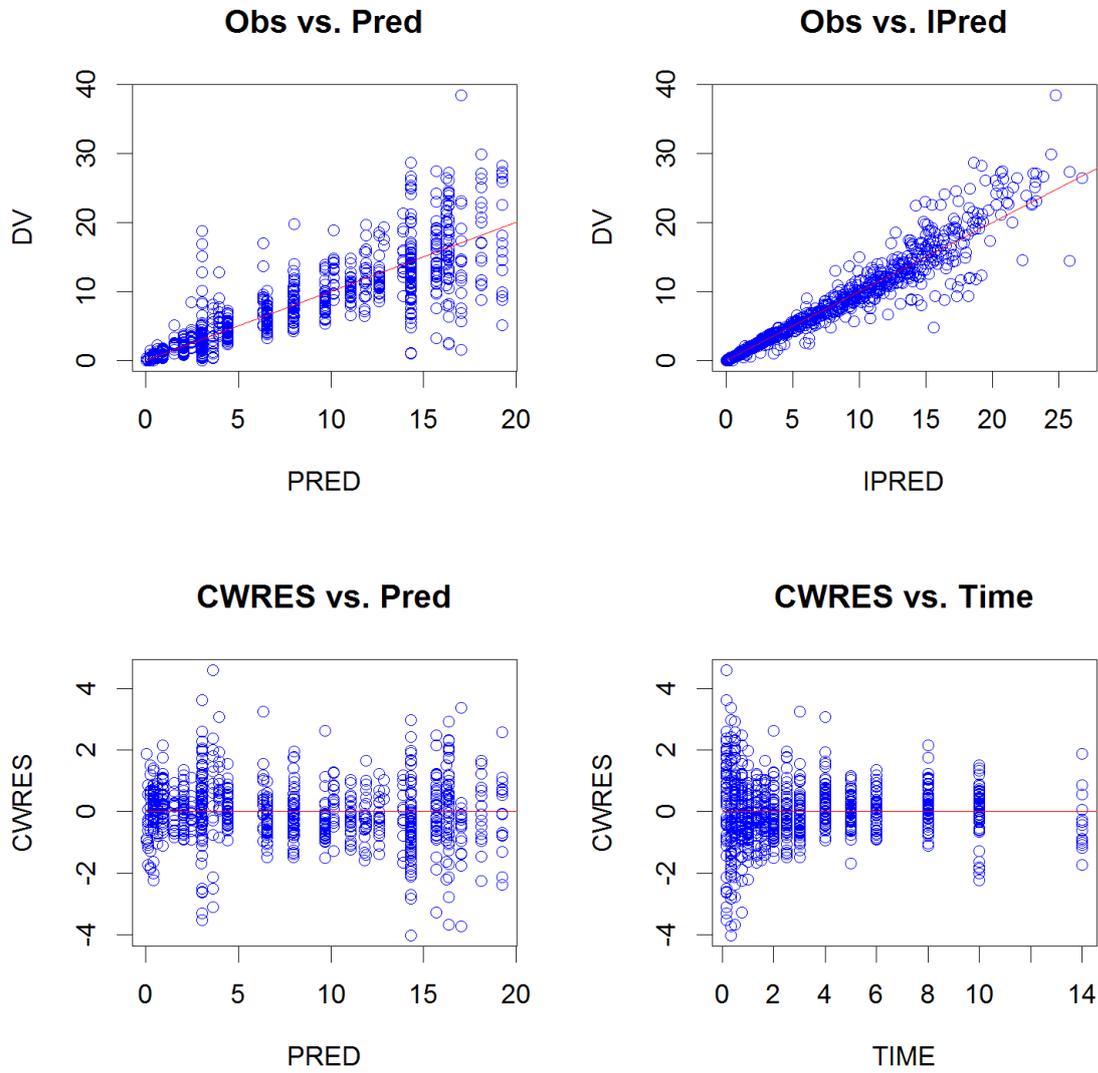


Figure 1S: Goodness-of-fit Plots. Upper panels show the observed data (Obs) vs. population predictions (Pred) (left) and the observed data vs. individual predictions (IPred) (right). Lower panels show the conditional weighted residuals (CWRES) vs. population predictions (left) and the CWRES vs time (left).

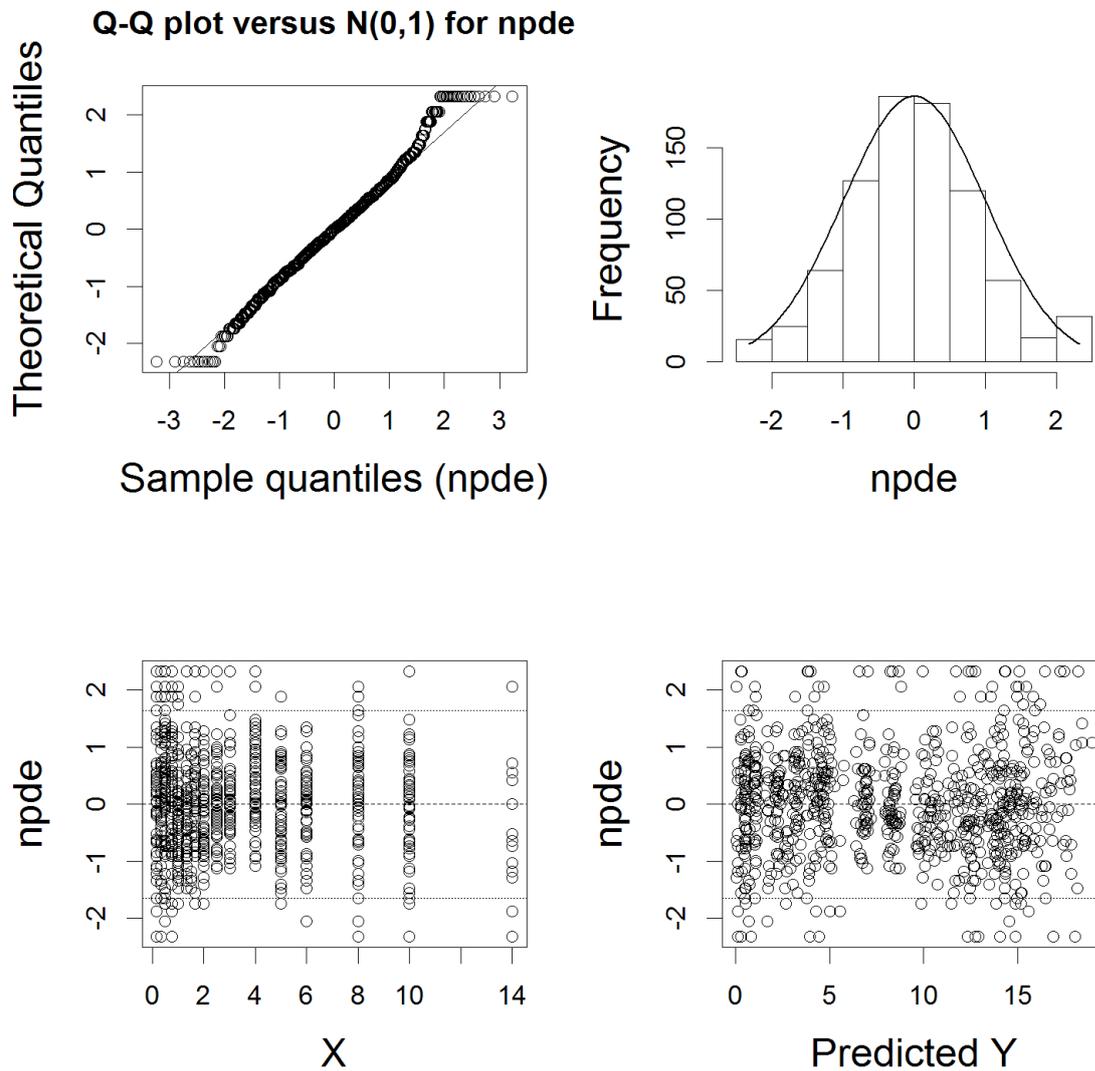


Figure 2S: Model validation: normalised prediction distribution errors. Upper panels show the QQ-plot of the distribution of the NPDEs for a theoretical $N(0, 1)$ distribution (left) and the histogram of the distribution of the NPDE together with the density of the standard normal distribution (right). Lower panels show the NPDEs vs. time (left) and NPDEs vs. individual predictions (right).

CHAPTER 4

Sampling optimisation in pharmacokinetic bridging studies: use of deferiprone in children with β -thalassaemia

Francesco Bellanti, Vincenzo Luca Di Iorio, Meindert Danhof, and Oscar Della Pasqua

Submitted for publication

Summary

Despite a wide experience with deferiprone, the optimum dosage in children aged less than 6 years remains to be established. This analysis is aimed at optimizing the design of a prospective clinical investigation for the evaluation of deferiprone pharmacokinetics in children. A one-compartment model with first order oral absorption was used for the design optimization. Different sampling schemes were evaluated under the assumption of a constrained population size. A sampling scheme with 5 samples post-dose per subject was found to be sufficient to ensure accurate characterization of the pharmacokinetics of deferiprone. Whereas the accuracy of parameters estimates was high, precision was slightly reduced due to the small sample size (> 30% for V_d/F and KA). AUC values (mean and SD) were found to be 33.37 (19.24) and 35.61 (20.22) $\mu\text{g/ml.h}$ and C_{max} values 10.17 (6.05) and 10.94 (6.68) $\mu\text{g/ml}$ in sparse and frequent sampling respectively. The results illustrate how ED-optimality concepts can be used to support PK bridging. Predefined sampling schemes and sample sizes do not warrant accurate model structure and parameter identifiability. Of importance is the accurate estimation of the magnitude of the covariate effects, as they may determine the dose recommendation for the population of interest.

4.1 Introduction

Patients with β -thalassaemia and other transfusion-dependent diseases develop iron overload from chronic blood transfusions and require regular iron chelation to prevent potentially fatal iron-related complications¹⁻⁵. Deferiprone (DFP) is the most extensively studied oral iron chelator to date. It has been authorized in Europe in 1999 for the treatment of iron overload in patients with β -thalassaemia major when deferoxamine (DFO) is contraindicated or inadequate. Despite a wide experience of DFP there are limited experimental data available on DFP in children and no data in children under 6 years of age.

Clinical studies, mostly in patients with beta-thalassaemia, have demonstrated that deferiprone at 75 to 100 mg/kg/day is capable of reducing iron burden in regularly transfused iron-overloaded patients⁵⁻¹⁰. The degree of iron loading is directly related to the level of iron intake from transfusions. Iron excretion with DFP, like with any chelator, was found to be dose-related. However, factors affecting response to deferiprone appear to include the degree of iron overload and duration, dosage and compliance with therapy. Although few long term comparative data are available, DFP at the recommended dosage of 75 mg/kg/day appears to be non-inferior to deferoxamine in the adult population. However, compliance is superior with DFP^{5,11}.

The optimum dosage of DFP in children less than 6 years of age remains to be established. Given that dose adjustment may be required in children, the aforementioned findings highlight the need for optimizing the dosing regimen and gathering supporting evidence for the dose rationale for subsequent assessment of efficacy in the pediatric population.

The information available so far in the adult population can be used to integrate the lack of knowledge in the pediatric population. The E11 guideline of the International Conference on Harmonization (ICH) supports the use of PK bridging concepts for the development of drugs in the pediatric population. Nevertheless, bridging studies can be implemented only if the following criteria are met: in the populations of interest the medicinal product should have the same indication, the disease process should be similar and the outcome of therapy should be comparable¹². This is true and applicable to patients affected by β -thalassaemia or other transfusion-dependent diseases.

Practical and ethical constraints impose special requirements for clinical trials in children¹²⁻¹⁵. The application of population pharmacokinetic (PK) analysis and PK bridging to sparse data allows reducing the burden in such a vulnerable population; yet it is important to optimize the quality of the information gathered.

The quality of the study can be dramatically improved through design optimization analysis. However, a PK model is needed to apply this methodology. When extrapolating information

from adults to children we have to make use of a hypothetical model which is derived from a different population than the population of interest. ED-optimality concepts can be applied to handle the uncertainty during the optimization procedure. Several studies have already shown how ED-optimization can be successfully applied to the design of clinical studies in children when extrapolating information from the adult population ¹⁶⁻²¹.

Based on simulation scenarios that take into account the impact of developmental growth, the aim of this analysis is to optimize the sampling times for the evaluation of the pharmacokinetics of deferiprone in a prospective clinical investigation in children younger than 6 years of age. The results of this trial will be subsequently used to define the most appropriate dosing regimen for this population.

4.2 Methods

Prospective Clinical Study: Design

A prospective study has been proposed to establish the pharmacokinetics of deferiprone in children. This will be investigated in a prospective multi-centre, randomized, single blind, and single dose study in patients affected by transfusion dependent hemoglobinopathies aged less than 6 years. Sample size of the study will range between a minimum of 18 up to a maximum of 30 evaluable pediatric patients. Patients will be randomized to three dose levels (8.3, 16.7 and 33.3 mg/kg) and will be exposed to a single dose of deferiprone. A maximum of 5 samples will be collected per patient. An optimization algorithm will be applied to evaluate the best sampling times in order to ensure high precision in parameter estimates and PK model identifiability.

Sampling Times Optimization

Several actions have been taken throughout the analysis that can be summarized in 6 major steps as depicted in the following flow-chart (Figure 1). Each step will be briefly discussed in the following paragraphs.

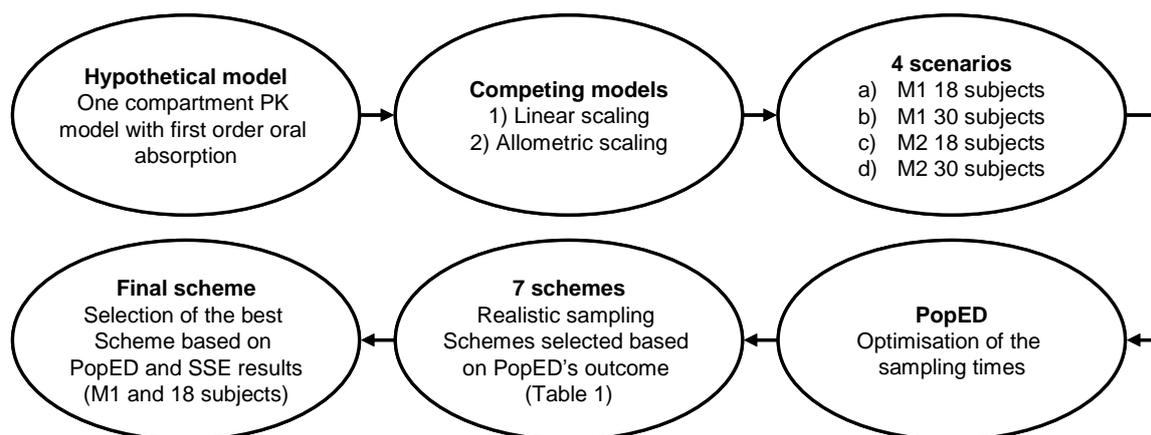


Figure 1. Data analysis flowchart.

Hypothetical PK Model

A one-compartment PK model with first-order absorption, lag-time to central compartment, and first-order elimination was used for the optimization of the sampling scheme ²². Between-subject variability (BSV) was estimated for apparent clearance (CL/F), apparent volume of distribution (V/F), and absorption rate constant (Ka). Residual variability was characterized by a proportional error model with a weighting factor.

Competing Models: extrapolation to the pediatric population

The reference model was previously developed by our group to explore DFP exposure in adults ²². With the purpose of optimizing the design of a prospective pediatric trial, the original model has been modified with the inclusion of two different covariate models, namely M1 (body weight linearly correlated with CL/F and V/F), and M2 (fixed allometric scaling: exponent of 0.75 on CL/F and 1.00 on V/F), in order to extrapolate deferiprone exposure to children.

Bearing in mind the objective of extrapolation across populations, focus was given to the model validation steps, which yield information about the variance structure and variance-covariance matrix. Visual predictive check (VPC) and NPDE summaries have been used to validate the model and to assess the suitability for simulation purposes. The software NONMEM (non-linear mixed effect modeling; release version 7.2.0) has been used for the procedure.

M1: the inclusion of body weight as a covariate on CL/F and V/F according to linear models was found to give the highest improvement in model performance in the previous investigation. The covariates were not included due to increase in model instability. In this case, given the different objective, we have considered including weight on CL/F and V/F into the final covariate model.

M2: Since the purpose of the analysis is to extrapolate information to the pediatric population, the use of allometric scaling (one of the current standard approach to extrapolate across populations) has been considered to evaluate possible discrepancies in optimizing the sampling schedule with two different approaches. Model parameters have been re-estimated with the new covariate relationships and the model has been tested for simulation purposes.

Diagnostic criteria, such as visual predictive checks and NPDE summaries have been used to assess model performance before the optimization of the study design (ED optimality) would be implemented. For both M1 and M2, Visual predictive check (VPC) plots indicate that model is not biased and is suitable for simulation purposes. In addition, NPDE summaries indicate that the discrepancy between predicted and observed values can be assumed to be normally distributed. VPC and NPDE for both models are provided in Figures S2 to S5 (see Appendix).

Optimization steps (and criteria)

The two hypothetical models have been used to identify the optimal sampling schedule in children after single dose deferiprone. The software for population experimental design “PopED” (release version 2.12) has been used to optimize sampling times and to assess precision in parameter estimates by evaluating the coefficient of variation (CV) for each parameter^{23–27}. Subsequently, the software NONMEM (non-linear mixed effect modeling; release version 7.2.0), has been used after sampling times optimization in order to assess model stability, and accuracy (RE: relative error) and bias (SME: standard mean error) in parameters estimates.

The following 4 scenarios have been evaluated in PopED to account for possible discrepancies between the two covariate models and differences in sample size:

- a) M1 with 18 subjects;
- b) M1 with 30 subjects;
- c) M2 with 18 subjects; and
- d) M2 with 30 subjects.

Sampling times have been optimized in the 4 scenarios for a simulated pediatric population which presented the following demographic characteristics: 50% males and females, and mean body weight of 20.5 kg (SD: 5.4). Subjects have been randomized to 3 dose levels as in the study design described above.

Information gathered through the optimization of the sampling times in PopED has then been used to create seven new realistic scenarios (each consisting of 3 sampling schemes) as

a result of a compromise between full optimization and feasibility in a real clinical trial. The seven scenarios have been compared and evaluated in order to define the final sampling schedule for the PK study. Furthermore, an extra scenario, consisting of an empirical, non-optimized sampling scheme has been evaluated along with the previous seven.

Given that no significant differences have been observed between the original four scenarios in PopED, scenario “a” (M1 with 18 subjects) has been selected and used for the final optimization step; this allowed also reducing significantly the computational effort of the analysis.

4.3 Results

Sampling times optimization in PopED

The results of the optimization of the sampling times is summarized in Figure 2, where the actual sampling times obtained in the 4 scenarios (“a”, “b”, “c”, “d”) are plotted. Each bar represents a sampling time selected during the optimization procedure, whereas each color indicates the contribution (in percentage) of the different scenarios.

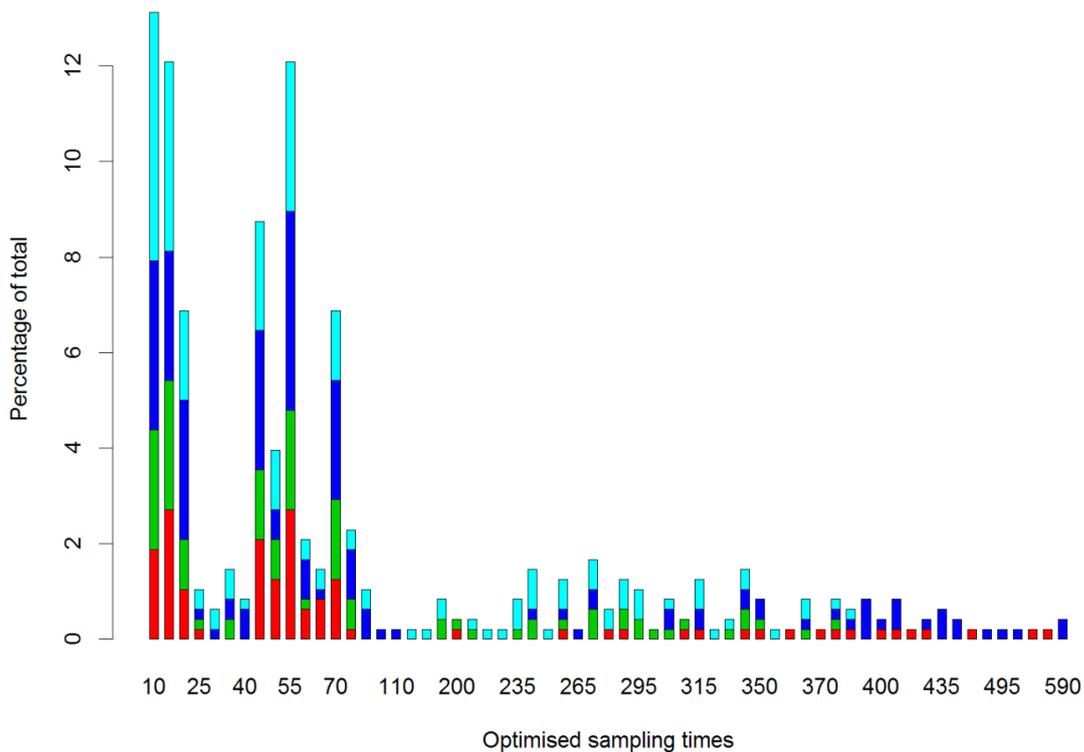


Figure 2. Sampling times obtained by ED-optimality. Sampling times selected during the optimization steps using ED-optimality: in red, green, dark blue and light blue are shown the time selected for scenarios a, b, c and d respectively. Percent of total indicates the percentage of cases for each set of optimized sampling times generated by PopED.

The data suggest that, independently of model and number of subjects, approximately 92% of sampling times should be collected within three intervals, namely:

- time window A: 32% in the range 10 to 20 minutes;
- time window B: 37% between 40 and 80 minutes;
- time window C: 23.75% after 200 minutes;

Based on the information collected in the previous step, and bearing in mind the compromise between full optimization and feasibility in a real clinical trial, seven sampling schemes (shown in Table 1 as scheme 1 to 7) have been generated and subsequently evaluated in PopED and NONMEM.

Table 1. Scenarios evaluated for sampling scheme selection.

Scenario	Sampling schemes	Scenario	Sampling schemes
0	Optimal Design	5	a: 10, 25, 50, 70, 360 b: 15, 45, 60, 270, 420 c: 20, 55, 75, 330, 480
1	a: 10, 25, 45, 70, 360 b: 15, 40, 60, 180, 420 c: 20, 55, 75, 240, 480	6	a: 10, 40, 65, 85, 360 b: 15, 45, 60, 270, 420 c: 20, 55, 75, 330, 480
2	a: 10, 30, 45, 180, 360 b: 15, 40, 60, 240, 420 c: 20, 50, 75, 300, 480	7	a: 10, 40, 65, 85, 360 b: 15, 50, 70, 270, 420 c: 20, 55, 75, 330, 480
3	a: 10, 40, 70, 180, 360 b: 15, 50, 80, 240, 420 c: 20, 60, 90, 300, 480	8*	30, 60, 120, 240, 480
4	a: 10, 25, 50, 75, 330 b: 15, 45, 60, 240, 360 c: 20, 55, 75, 270, 420	/	/

* Empirical sampling scheme reflecting the current practice, i.e., non-optimised design.

Evaluation of seven realistic sampling schemes

As previously mentioned in the methods section, given that no major differences have been observed in the previous step between the two different models (M1 and M2) and different number of subjects (18 vs. 30), only model M1 with a total number of 18 subjects was used for the second part of the optimization.

Table S1 (see Appendix) shows the coefficient of variation (CV) for the different scenarios compared to the optimal sampling scheme. No major differences can be observed between the different scenarios, except for the non-optimized scheme (number 8) in which a remarkably higher uncertainty for the following parameters can be observed: CL slope, V

slope, and Ka. Furthermore, results clearly highlight the poor performance of the model if an empirical (non-optimized) pharmacokinetic sampling scheme is used.

Table S2 (see Appendix) shows model stability results based on NONMEM stochastic simulation and estimation (SSE). In this overview, schemes 3 and 7 show higher stability, as compared to the other sampling schemes. On the other hand, scheme 8 shows the worst result out of the 9 scenarios. Finally, Figures 3, 4 and S1 (for Figure S1 see Appendix) show measures of accuracy (RE) and bias (SME) for the main parameters of interest for the different sampling schemes.

Altogether, scheme 7 was the closest one to the fully optimized sampling scheme, providing the best combination of results in terms of bias (SME) and accuracy (RE) of parameters estimates. This was also true in terms of model robustness, with only 1 failed minimization and 435 successful covariate steps out of 500 runs (Table S2; see Appendix).

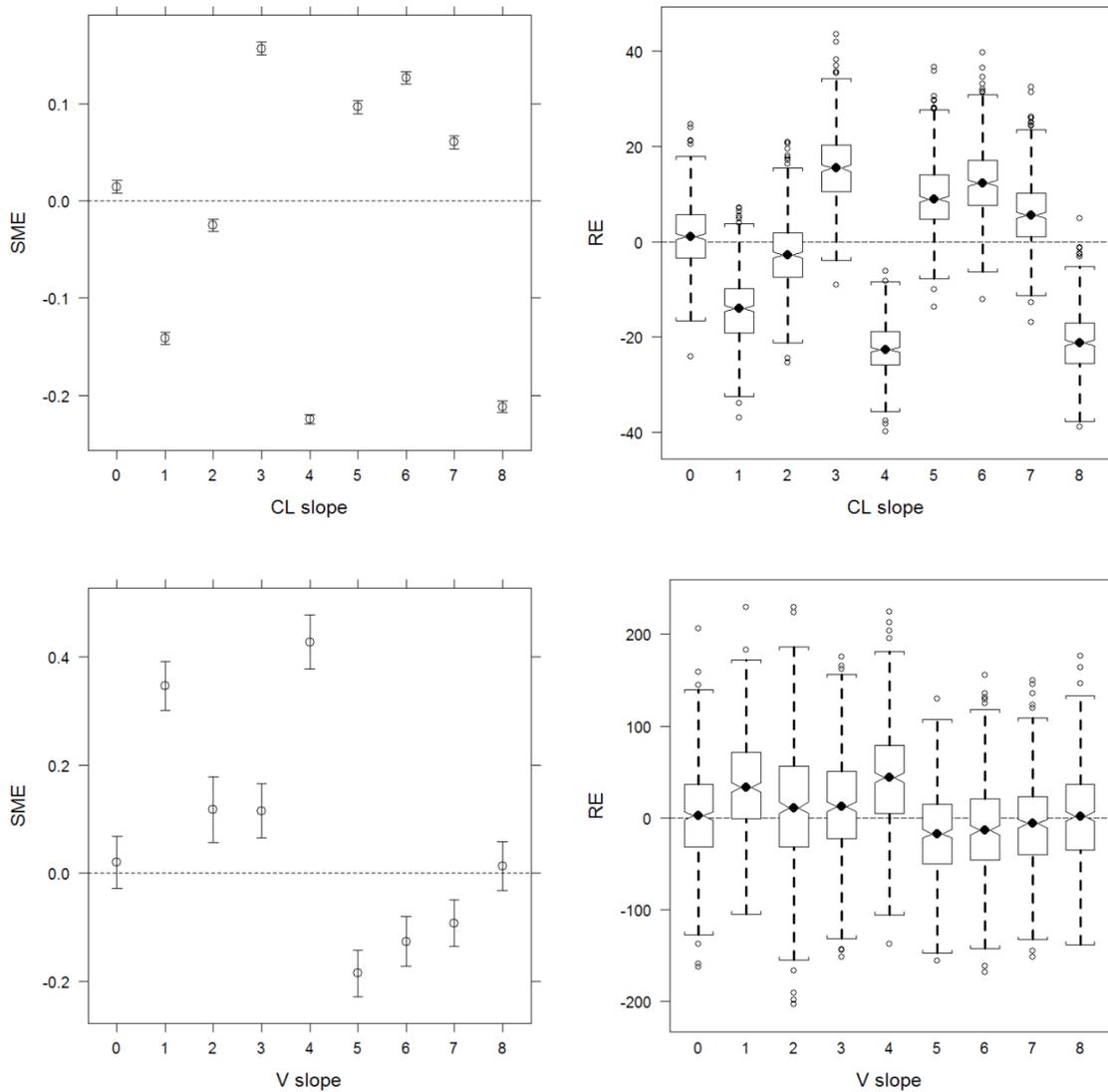


Figure 3. SME and RE for the slope parameters describing the effect of body weight on clearance and volume of distribution. Standard Mean Error (SME) and Relative Error (RE) estimates indicate, respectively, the bias and accuracy in parameter estimates for the different sampling schemes. Top panels: summary for the slope parameter describing the effect of body weight on clearance. Bottom panels: summary for the slope parameter describing the effect of body weight on volume of distribution.

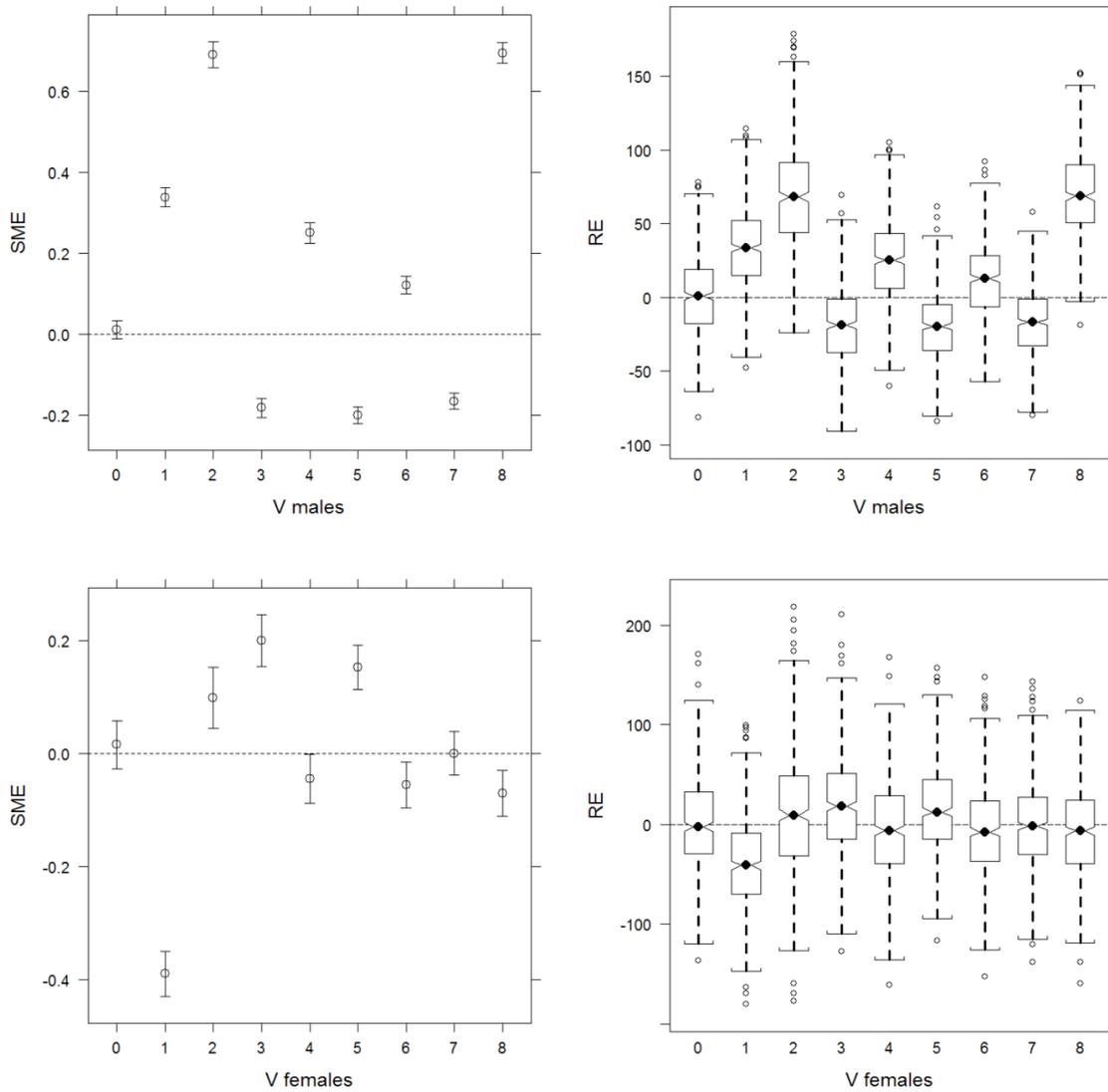


Figure 4. SME and RE for the volume of distribution in males and females. Standard Mean Error (SME) and Relative Error (RE) estimates indicate, respectively, the bias and accuracy in parameter estimates for the different sampling schemes. Top panels: summary for volume of distribution in males Bottom panels: summary for volume of distribution in females.

Secondary PK parameters: sparse sampling vs. rich sampling

To further assess the suitability of scheme seven for the prospective PK study, focus was given to the ability of the model to predict secondary PK parameters. Model-predicted AUC and Cmax based on the sparse sampling scheme were compared with estimates obtained according to a rich sampling scheme (12 samples per subject). AUC has been calculated with the trapezoidal rule.

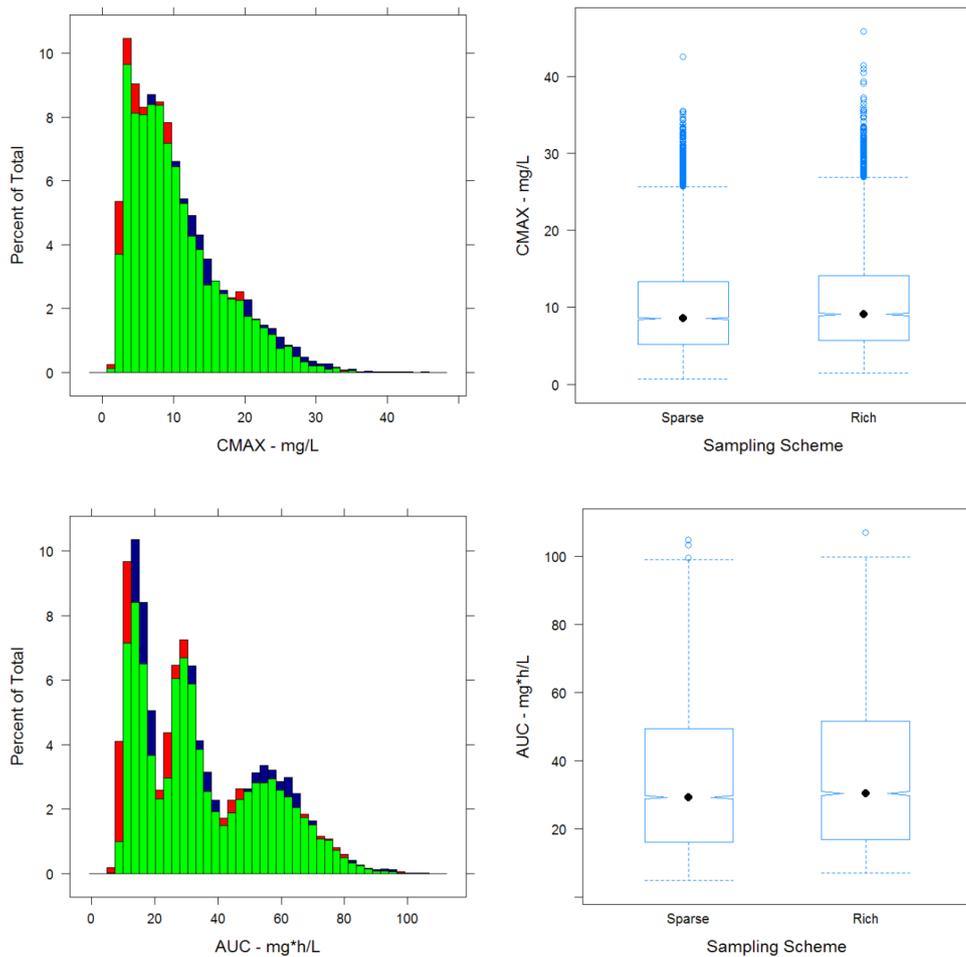


Figure 5. AUC and CMAX estimation using sparse (5) vs. frequent (12) sampling. Model predicted deferiprone systemic exposure expressed as AUC and Cmax. The final scheme selected during the optimization procedure (scheme 7) with sparse sampling (5 samples) is compared with rich sampling (12 samples). Top panels: histogram and boxplot describing the distribution of Cmax. Bottom panels: histogram of the distribution of AUC. Percent of total represents the percentage of cases for each set of 500 simulations with 18 patients in each simulated trial. Red: sparse sampling; Blue: rich sampling; Green: overlapping area.

As shown by the distributions and box-plots of AUC and Cmax (Figure 5), no significant differences were detected for sparse and frequent sampling. AUC values (mean and standard deviation) were found to be 33.37 (19.24) and 35.61 (20.22) $\mu\text{g/ml.h}$ and Cmax values 10.17 (6.05) and 10.94 (6.68) $\mu\text{g/ml}$ in sparse and frequent sampling respectively.

4.4 Discussion and Conclusion

Clinical trials in children represent a very challenging phase in drug development. Given the ethical and practical constraints imposed to experimental protocols in this vulnerable population^{12–15}, the information gathered per subject becomes significantly more important, as compared to the adult population. In addition, increasing evidence suggests the lack of suitability of empirical protocols in pediatric research. This limitation has therefore prompted clinical scientists and drug developers to consider the use of alternative approaches for the evaluation of pharmacokinetics, efficacy and safety in children^{28–34}.

Our results show that highly informative data can be generated whilst reducing the burden of clinical trials in children. From a methodological perspective, these data also reinforce the benefits associated with the use of ED-optimality concepts for the design of pediatric clinical studies^{16–21}. In fact, it should be noted that whereas optimal design is normally based on a model representative of the population of interest, our analysis was aimed at an extrapolation model derived from data available in adult patients, i.e., we have used a hypothetical model. Consequently, the optimization procedures carry a certain degree of uncertainty.

Nevertheless, the major advantage of using ED-optimization is that this methodology accounts for the uncertainty parameter estimates and in the effect of covariates during the optimization procedure. More specifically, two scenarios with 20% and 40% uncertainty on the main parameters of interest (clearance and volume of distribution) were evaluated. The data shown throughout this report reflect the first case only. Increased uncertainty in parameter distributions had no significant impact on the optimization and selection of sampling times. Lastly, to ensure a comprehensive evaluation of the assumptions underlying the nature and magnitude of the covariate effects on the systemic exposure in children, we have resorted to two competing models (M1 and M2). These models enabled the identification of the best sampling scheme for the prospective pharmacokinetic study taking into account potential differences in the disposition of deferiprone in children younger than 6 years of age.

Based on our experience in pediatric clinical pharmacology, fully optimized designs (i.e., individually optimized) are not realistically applicable to pediatric trials. Our objective in this regard was to identify a final sampling scheme that resulted from a compromise between

full optimization and feasibility in a real setting. The three time windows selected during the optimization procedures were found to be independent of model specification and sample size. Moreover, they reflect the requirements for estimating specific pharmacokinetic parameters, i.e., lag-time and K_a (time window A), V/F (time window B), and CL/F (time window C) respectively.

Finally, it should be noted that our analysis clearly highlights the poor performance of an empirical (i.e., non-optimized) sampling scheme (scheme 8), especially when dealing with sparse sampling. By contrast, scenario 7 showed the best option in terms of coefficient of variation, relative errors and standard mean errors as well as for what concerns model stability. In addition, it allowed for correct predictions of AUC and C_{max} .

Potential limitations

The final decision on the sampling scheme to be used in the prospective study had to take another unknown factor into account, namely the uncertainty around the time at which the peak concentration occurs (T_{max}). Keeping in mind that this exercise is based on a model derived from pharmacokinetic data in adult patients, there might be some differences in children below 6 years of age. Such differences may occur despite the quick absorption after administration of the drug as an oral solution. They may also be caused by difficulties in the administration of the drug to the very young children. We have therefore included these considerations in scenarios 3, 6 and 7, but there are no data available at the moment to confirm the assumptions regarding the possible shift of T_{max} .

In conclusion, our analysis illustrates and confirms that despite feasibility issues, ED-optimality concepts can be used to optimize study design, particularly with regard to the pediatric population. Predefined sampling schemes and sample sizes do not warrant accurate model structure and parameter identifiability. In addition, it shows that the optimization of study design does not require necessarily the use of the final model for the population of interest; the combination between ED-optimization and the information carried by a hypothetical model is sufficient to significantly increase the quality of the information collected in a prospective clinical trial. Nevertheless, remains of particular importance the accurate estimation of the magnitude of the covariate effects, as they may determine the final dose recommendation for the population of interest.

References

1. R.Gibbons, D.R.Higgs, J.M.Old, Olivieri Nancy F., Thein Swee Lay WGW. The Thalassaemia Syndromes - Fourth Edition. *Blackwell Sci.* 2001.
2. Borgna-Pignatti C, Cappellini MD, De Stefano P, et al. Survival and complications in thalassemia. *Ann N Y Acad Sci.* 2005; 1054:40-7.
3. Borgna-Pignatti C, Cappellini MD, De Stefano P, et al. Cardiac morbidity and mortality in deferoxamine- or deferiprone-treated patients with thalassemia major. *Blood.* 2006;107(9):3733-7.
4. Cunningham MJ, Macklin EA Neufeld EJ, Cohen AR. Complications of beta-thalassemia major in North America. *Blood.* 2004; 104(1):34-9.
5. Galanello R, Campus S. Deferiprone chelation therapy for thalassemia major. *Acta Haematol.* 2009; 122(2-3):155-64.
6. Piga A, Roggero S, Salussolia I, Massano D, Serra M, Longo F. Deferiprone. *Ann N Y Acad Sci.* 2010; 1202:75-8.
7. Barman Balfour JA, Foster RH. Deferiprone: a review of its clinical potential in iron overload in beta-thalassaemia major and other transfusion-dependent diseases. *Drugs.* 1999; 58(3):553-78.
8. Won SC, Han DK, Seo JJ, et al. Efficacy and safety of deferiprone (Ferriprox), an oral iron-chelating agent, in pediatric patients. *Korean J Hematol.* 2010; 45(1):58-61.
9. ElAlfy MS, El Alfy M, Sari TT, Lee CL, Tricta F, El-Beshlawy A. The safety, tolerability, and efficacy of a liquid formulation of deferiprone in young children with transfusional iron overload. *J Pediatr Hematol Oncol.* 2010; 32(8):601-5..
10. Diav-citrin O. Oral iron chelation with deferiprone. *New Front Pediatr Drug Ther.* 1997;44(1):235-247.
11. Maggio A, D'Amico G, Morabito A, et al. Deferiprone versus Deferoxamine in Patients with Thalassaemia Major: A Randomized Clinical Trial. *Blood Cells, Mol Dis.* 2002;28(2):196-208.
12. ICH Topic E11. http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002926.pdf. 2001.
13. EMA. Paediatric Workshops - http://www.emea.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000416.jsp&mid=WC0b01ac0580118a19.
14. Henschel AD, Rothenberger LG, Boos J. Randomized clinical trials in children--ethical and methodological issues. *Curr Pharm Des.* 2010; 16(22):2407-15.

15. Sammons HM, Starkey ES. Ethical issues of clinical trials in children. *Paediatr Child Health (Oxford)*. 2012; 22(2):47-50.
16. Ogungbenro K, Aarons L. How many subjects are necessary for population pharmacokinetic experiments? Confidence interval approach. *Eur J Clin Pharmacol*. 2008; 64(7):705-13.
17. Bouillon-Pichault M, Jullien V, Bazzoli C, Pons G, Tod M. Pharmacokinetic design optimization in children and estimation of maturation parameters: example of cytochrome P450 3A4. *J Pharmacokinet Pharmacodyn*. 2011;38(1):25-40.
18. Ogungbenro K, Matthews I, Looby M, Kaiser G, Graham G, Aarons L. Population pharmacokinetics and optimal design of paediatric studies for famciclovir. *Br J Clin Pharmacol*. 2009; 68(4):546-60.
19. Nyberg J, Karlsson MO, Hooker AC. Simultaneous optimal experimental design on dose and sample times. *J Pharmacokinet Pharmacodyn*. 2009; 36(2):125-45.
20. Hooker AC, Foracchia M, Dodds MG, Vicini P. An Evaluation of Population D-Optimal Designs Via Pharmacokinetic Simulations. *Ann Biomed Eng*. 2003; 31(1):98-111.
21. Hooker A, Vicini P. Simultaneous population optimal design for pharmacokinetic-pharmacodynamic experiments. *AAPS J*. 2005; 7(4):E759-85.
22. Bellanti F, Danhof M, Della Pasqua O. Population pharmacokinetics of deferiprone in healthy subjects. *Submitt Publ*.
23. 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.
24. Foracchia M, Hooker A, Vicini P, Ruggeri A. POPED, a software for optimal experiment design in population kinetics. *Comput Methods Programs Biomed*. 2004; 74(1):29-46.
25. Dodds MG, Hooker AC, Vicini P. Robust population pharmacokinetic experiment design. *J Pharmacokinet Pharmacodyn*. 2005; 32(1):33-64.
26. Mentre F, Mallet A, Baccar D. Optimal design in random-effects regression models. *Biometrika*. 1997; 84(2):429-442.
27. Retout S, Duffull S, Mentré F. Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs. *Comput Methods Programs Biomed*. 2001; 65(2):141-51.
28. Cella M, Gorter de Vries F, Burger D, Danhof M, Della Pasqua O. A model-based approach to dose selection in early pediatric development. *Clin Pharmacol Ther*. 2010; 87(3):294-302.

29. Bellanti F, Della Pasqua O. Modelling and simulation as research tools in paediatric drug development. *Eur J Clin Pharmacol*. 2011;67 Suppl 1:75-86.
30. Cella M, Zhao W, Jacqz-Aigrain E, Burger D, Danhof M, Della Pasqua O. Paediatric drug development: are population models predictive of pharmacokinetics across paediatric populations? *Br J Clin Pharmacol*. 2011; 72(3):454-64..
31. Manolis E, Osman TE, Herold R, et al. Role of modeling and simulation in pediatric investigation plans. *Paediatr Anaesth*. 2011; 21(3):214-21.
32. 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.
33. Knibbe CAJ, Krekels EHJ, Danhof M. Advances in paediatric pharmacokinetics. *Expert Opin Drug Metab Toxicol*. 2011; 7(1):1-8.
34. Anderson BJ, Allegaert K, Holford NHG. Population clinical pharmacology of children: general principles. *Eur J Pediatr*. 2006; 165(11):741-6.

Appendix

Table S1. Coefficient of variation for pharmacokinetic parameters of the model assuming a linear relationship between body weight and CL and Vd. Estimates calculated according to study design including 18 patients and 20% uncertainty in the parameter estimates for CL and V.

	CV (%)								
	SAMPLING SCHEMES								
PARAMETERS	0	1	2	3	4	5	6	7	8
CL intercept*	/	/	/	/	/	/	/	/	/
CL slope	7.4	7.2	6.9	6.9	7.0	7.0	7.0	7.0	29.4
BSV CL	59.7	57.3	53.9	53.2	55.2	55.4	55.0	54.8	52.7
V Males	22.5	23.4	24.1	24.1	23.6	23.4	23.2	23.3	23.8
V Females	43.8	44.6	45.8	45.8	44.9	44.5	44.2	44.3	45.5
V slope	48.9	49.7	50.8	51.1	49.9	49.6	49.4	49.6	77.4
BSV V	57.8	59.5	61.6	62.8	60.3	59.5	59.0	59.4	59.9
Ka	26.9	29.7	33.3	37.3	29.7	29.5	37.0	38.1	379.2
BSV Ka	37.4	39.2	40.1	40.4	39.2	39.2	40.4	40.4	72.2

*The parameter describing the intercept for the linear function between body weight and clearance has been fixed during the analysis. Therefore no CV values are reported.

Table S2. Evaluation of model stability and parameter identifiability based on different sampling schemes. Values represent the results of 500 runs.

SAMPLING SCHEME	Minimisation successful	Covariate Step Successful	Estimate near boundary
0	496	391	108
1	409	306	118
2	476	398	82
3	490	443	47
4	454	362	101
5	446	419	28
6	471	419	33
7	499	435	63
8	283	217	100

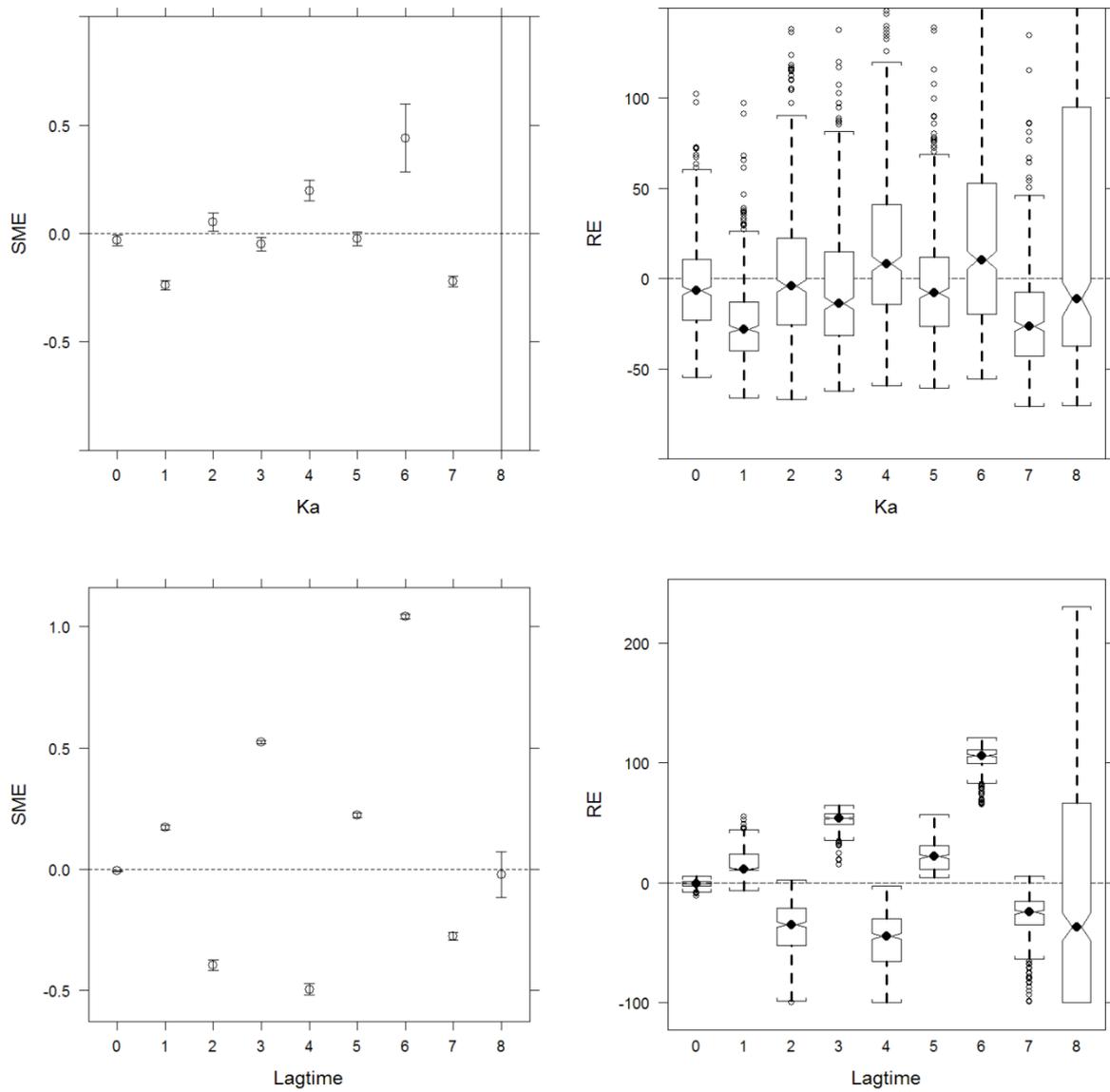


Figure S1. SME and RE for the slope parameters describing the effect of body weight on clearance and volume of distribution. Standard Mean Error (SME) and Relative Error (RE) estimates indicate, respectively, the bias and accuracy in parameter estimates for the different sampling schemes. Top panels: summary for absorption rate constant (Ka). Bottom panels: summary for lag time.

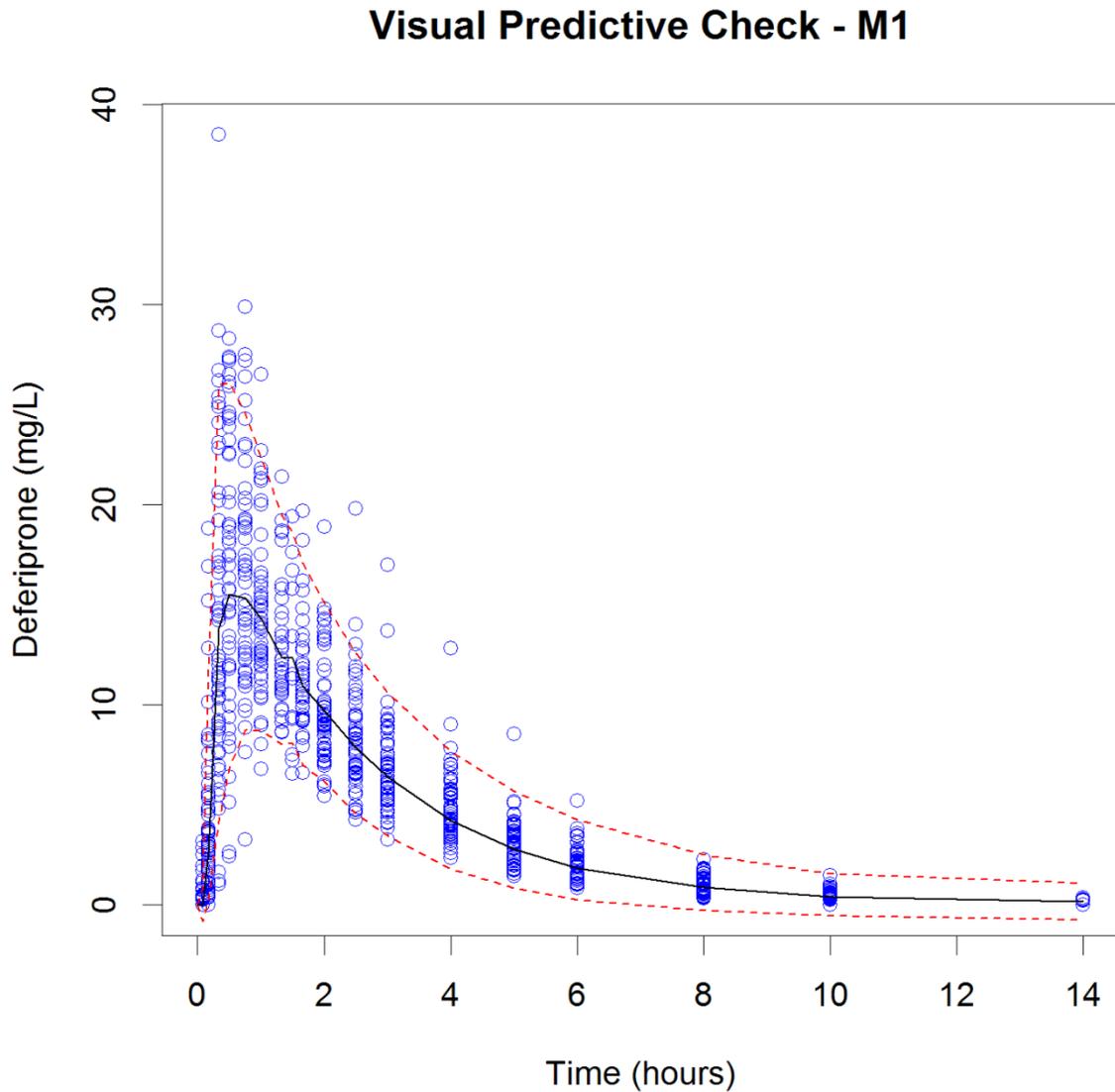


Figure S2. Visual Predictive Check for model M1. Simulated concentration vs. time course profile of deferiprone according to a pharmacokinetic model in which weight is correlated with clearance and volume according to a linear function. Observed data are plotted using open circles; the solid line represents the median of the simulated data; the dashed lines represent the 5th and 95th percentiles of the simulated data.

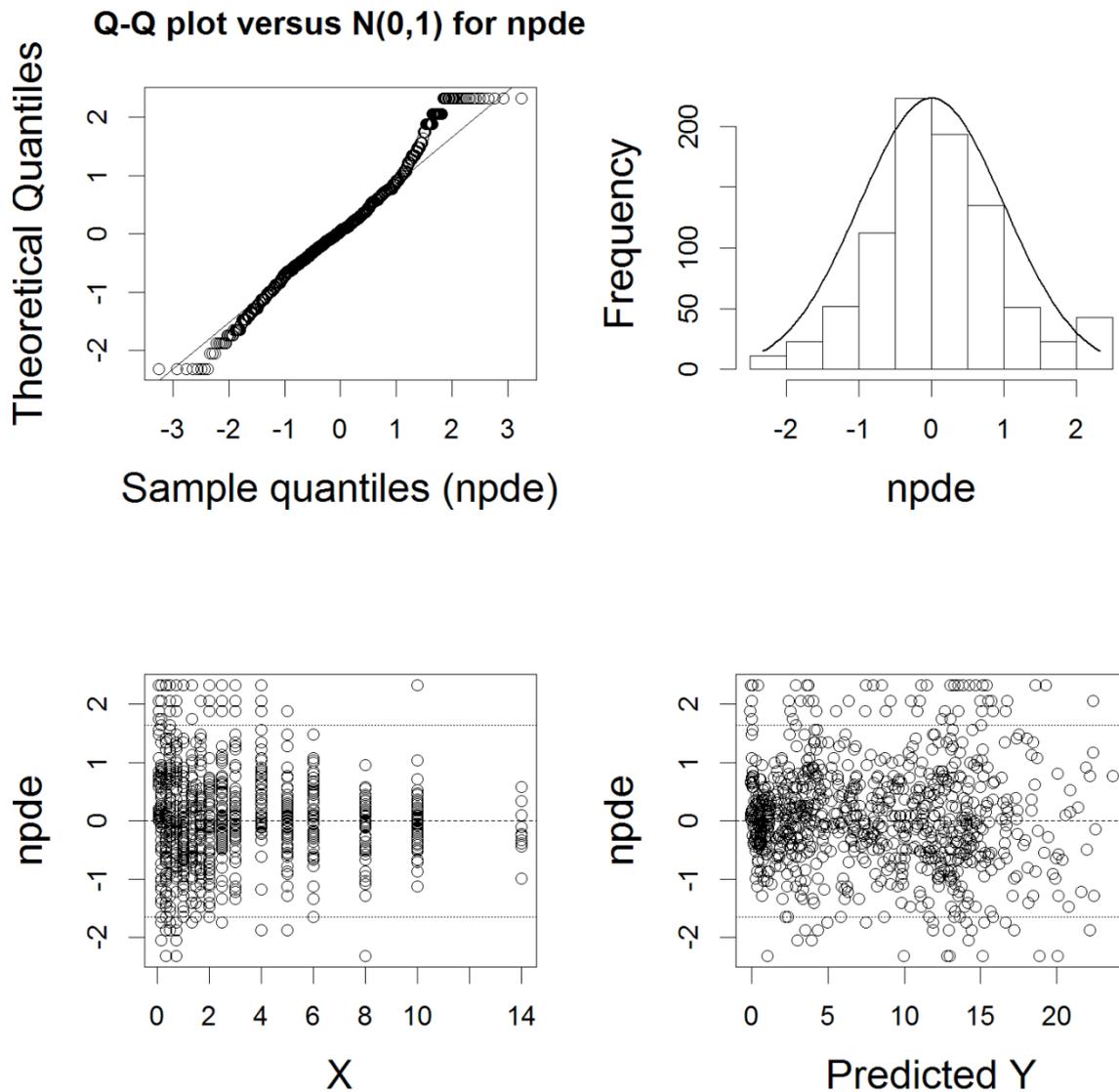


Figure S3. Normalised prediction distribution errors (NPDE) for model M1. Given that the accuracy of model predictions also depends on the variance structure, special attention was paid to the evaluation of model misspecifications for the random effects. The normalised prediction distribution errors (NPDE) method was applied for an in-depth diagnosis of potentially poor behaviour. Top left: QQ-plot of the distribution of the NPDE versus the theoretical N (0,1) distribution; Top right: Histogram of the distribution of the NPDE, with the density of the standard Gaussian distribution overlaid; Bottom left: NPDE versus time and Bottom right: NPDE versus PRED.

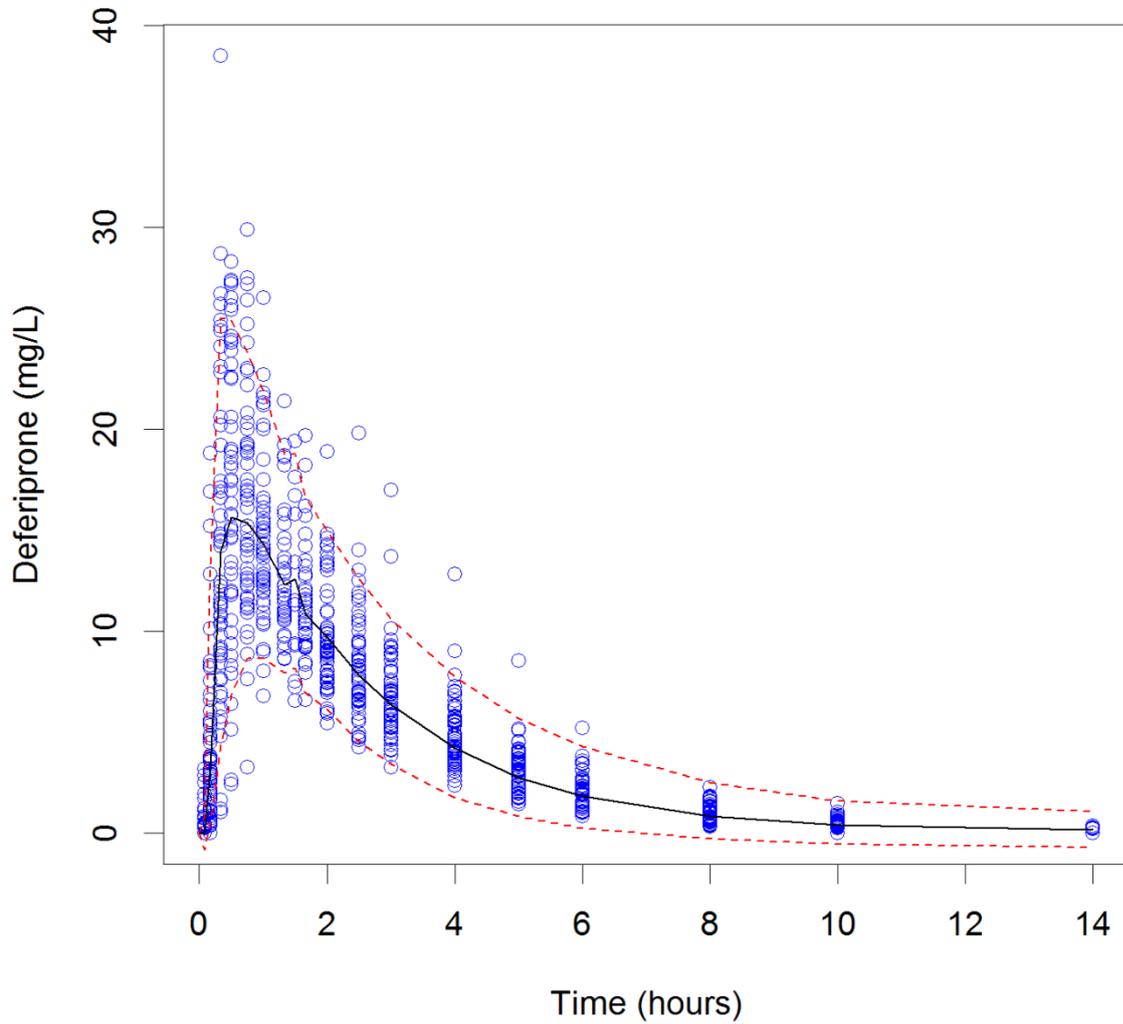
Visual Predictive Check - M2

Figure S4. Visual Predictive Check for model M2. Simulated concentration vs. time course profile of deferiprone according to a pharmacokinetic model in which weight is correlated with clearance and volume according to an allometric function. Observed data are plotted using open circles; the solid line represents the median of the simulated data; the dashed lines represent the 5th and 95th percentiles of the simulated data.

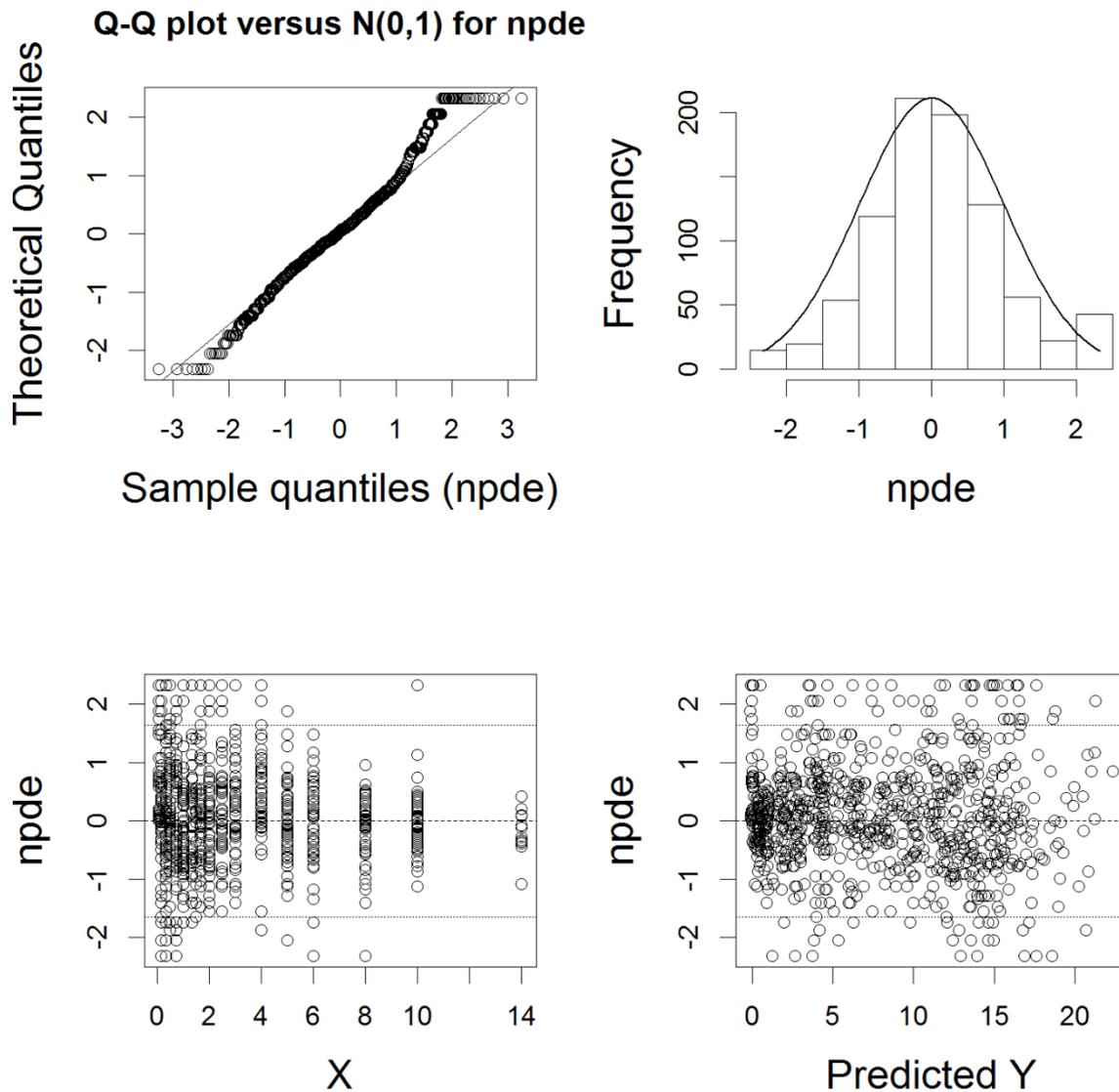


Figure S5. Normalised prediction distribution errors (NPDE) for model M2. Given that the accuracy of model predictions also depends on the variance structure, special attention was paid to the evaluation of model misspecifications for the random effects. The normalised prediction distribution errors (NPDE) method was applied for an in-depth diagnosis of potentially poor behaviour. Top left: QQ-plot of the distribution of the NPDE versus the theoretical N (0,1) distribution; Top right: Histogram of the distribution of the NPDE, with the density of the standard Gaussian distribution overlaid; Bottom left: NPDE versus time and Bottom right: NPDE versus PRED.

CHAPTER 5

Model-based dosing recommendations for the use of deferiprone in children affected by transfusional iron overload younger than 6 years of age

F. Bellanti, G.C. Del Vecchio, M.C. Putti, A. Maggio, A. Filosa, C. Cosmi, L. Mangiarini, M. Spino, J. Connelly, A. Ceci and O. Della Pasqua

Submitted for publication

Summary

Despite long clinical experience with deferiprone, there is still limited information on its pharmacokinetics in children and essentially none in children below 6 years of age. The objective of this analysis is to characterise the pharmacokinetics of deferiprone in the target population using a model-based approach and to assess the effect of demographic and physiological factors on drug exposure. Furthermore, it is our aim to ascertain whether equivalent doses on a mg/kg basis produce PK in children consistent with that in adults. Data from 18 paediatric patients receiving deferiprone orally (solution 80 mg/ml) were used for model building purposes. A one-compartment model with first order oral absorption was found to best describe the pharmacokinetics of deferiprone. Goodness-of-fit plots, visual predictive check (VPC) and NPDE summaries indicated that the model provides an unbiased description of the data. Simulation scenarios revealed that similar mg/kg dose levels yield comparable exposure in children and adults, with median AUC values respectively of 340.6 and 318.5 $\mu\text{mol/L}\cdot\text{h}$ at 75 mg/kg/day and 453.7 and 424.2 at 100 mg/kg/day t.i.d. doses evenly spaced. Based on these findings, a dosing regimen of 25 mg/kg t.i.d. is recommended in children below 6 years of age, with the possibility of titration up to 33.3 mg/kg t.i.d.

5.1 Introduction

Patients with hemoglobinopathies and certain other conditions affecting the ability to synthesize haemoglobin may require life-long blood transfusion therapy to survive. This chronic intervention results in a series of potential complications, with iron overload being an inevitable consequence within a few years. Chelation therapy is therefore required to prevent potentially fatal iron-related complications¹⁻⁵. Deferiprone (DFP) is a hydroxypyridinone, which was authorised in Europe in 1999 for the treatment of iron overload in patients with β -thalassaemia major when deferoxamine (DFO) is contraindicated or inadequate. When administered orally, DFP is rapidly and well absorbed. Plasma levels show peak concentrations (C_{max}) within 1 hour of administration. Food reduces its absorption rate without affecting the overall exposure to the drug. In patients with β -thalassaemia, the administration of deferiprone at doses of 75 mg/kg/day as a twice-daily regimen yields C_{max} of 34.6 mg/L and area under the plasma concentration-time curve (AUC) of 137.5 mg/L • h^{6,7}. On the other hand, peak serum concentrations were 17.53 mg/L and 11.82 mg/L in fasting and fed states, respectively after a dose of 25 mg/kg⁸. DFP is for the most part inactivated by glucuronidation (>85%) and more than 90% of the drug is removed from plasma within 6 hours of ingestion, with an elimination half-life of 1 to 2.5 hours in patients affected by β -thalassaemia^{5,6,9-16}. DFP forms a 3:1 complex with iron, which is removed mainly through the kidneys, as is the free parent drug.

Despite the extensive clinical experience with DFP, there are few PK data in children, and effectively none in children under 6 years of age. To cover this gap Deferiprone was included in the list of priority prepared by the PDCO-EMA. The main objective of this analysis is to appropriately characterise the systemic exposure of DFP in paediatric patients aged less than 6 years using a model-based approach and to assess the effect of demographic and physiological factors on the drug's pharmacokinetics. Furthermore, it is our endeavour to identify the dose levels yielding DFP exposures comparable to those in adults.

5.2 Methods

Clinical Study

This experimental and modelling study is a multi-centre, randomised, single blind, single dose PK study to evaluate the pharmacokinetics of DFP in children aged from one month to less than 6 years affected by transfusion-dependent haemoglobinopathies.

The pharmacokinetics of deferiprone was evaluated using data collected from the clinical study: DEEP-1 PK Study (EudraCT, 2012-000658-67), in which paediatric patients affected by transfusion-dependent haemoglobinopathies received a single oral dose of DFP as an 80 mg/ml solution. Patients undergoing a chronic transfusion program (receiving at least 150 ml/kg/year of packed red blood cells) and, if naïve to any chelation therapy, having ferritin

levels above 800 ng/ml were considered eligible for the study. In addition, amongst other criteria, patients with Hb levels less than 8 g/dl, abnormal liver function, and severe heart dysfunction secondary to iron overload or serum creatinine levels above the upper normal level were not considered eligible for inclusion in the study. Patients were randomised to three dose levels: 8.3, 16.7 and 33.3 mg/kg. The study was performed within the DEEP Consortium (www.deep.cvbf.net) according to an approved PIP (EMA-001126-PIP01-10). The study protocol was approved by concerned Ethics Committees and all experimental procedures performed according to good clinical practice guidelines. In brief, 18 children aged from 1 month to less than 6 years (9 males and 9 females) who had received the active medication were included in the analysis. Recruitment of up to 30 patients was provided for by protocol to ensure a minimum sample size of 18 evaluable subjects. In practice, the use of nonlinear mixed-effects modelling allowed completing the study with the data of the first 18 evaluable subjects by providing accurate and precise estimates of the main parameters of interest. Blood samples for the evaluation of deferiprone concentrations were taken before (one pre-dose sample) and at the following sampling times after dosing: 0.167, 0.25, 0.333, 0.67, 0.83, 0.916, 1.083, 1.167, 1.25, 1.416, 4.5, 5.5, 6, 7 and 8 hours. A maximum of 5 post-dose samples were collected per subject according to 3 sampling schemes selected based on an optimal design analysis previously performed by our group (unpublished results). Blood samples were drawn by peripheral venous catheter following discard of 2 ml of blood; catheters were filled with saline (i.e., saline lock) between sampling times. Mean (sd) age (years), body weight (kg) and height (cm) of the patient population were 3.62 (1.33), 16.08 (3.18) and 98.95 (9.16) respectively.

Bioanalysis

Deferiprone plasma concentrations were analysed by the laboratory of the Division of Pharmacology (Leiden, the Netherlands) using a validated method previously developed by ApoPharma (Toronto, Canada) consisting of high performance liquid chromatography with UV detection (HPLC-UV). Extraction of deferiprone from supernatant was performed after precipitation of plasma proteins by trichloroacetic acid (TCA - 15%) and centrifugation at 10,000 g for 20 minutes at 4 °C. The analytical column used for the analysis was a Hamilton PRP-1 and separation of the chromatogram of interest was achieved using an isocratic mobile phase (pH 7.0). The analytical range was between 3.13 and 800 µM (equivalent to 0.43 to 111 µg/ml); and an R² value greater than 0.98 was required to accept the standard curve. The lower limit of quantification (LLOQ) was 0.238 µM (equivalent to 0.033 µg/ml). Inter- and Intra-day accuracy and precision were always below 6 %, except for the inter-day precision at 3.13 µM which was found to be 10.7 %.

Pharmacokinetic Modelling

Nonlinear mixed effects modelling was performed in NONMEM version 7.2 (Icon Development Solutions, USA). Model building criteria included: (i) successful minimisation, (ii) standard error of estimates, (iii) number of significant digits, (iv) termination of the covariance step, (v) correlation between model parameters and (vi) acceptable gradients at the last iteration.

Fixed and random effects were introduced into the model in a stepwise manner. Inter-individual variability in pharmacokinetic parameters was assumed to be log-normally distributed. A parameter value of an individual i (post hoc value) is therefore given by the following equation:

$$\theta_i = \theta_{TV} * e^{\eta_i}$$

in which θ_{TV} is the typical value of the parameter in the population and η_i is assumed to be a random variable with zero mean and variance ω^2 . Residual variability, which comprises measurement and model error, was described with a proportional error model. This means for the j th observed concentration of the i th individual, the relation Y_{ij} :

$$Y_{ij} = F_{ij} + \varepsilon_{ij} * W$$

where F_{ij} is the predicted concentration and ε_{ij} the random variable with mean zero and variance σ^2 . W is a proportional weighing factor for ε .

Goodness of fit was assessed by graphical methods, including population and individual predicted vs. observed concentrations, conditional weighted residual vs. observed concentrations and time, correlation matrix for fixed vs. random effects, correlation matrix between parameters and covariates and normalised predictive distribution error (NPDE)^{17,18}. Comparison of hierarchical models was based on the likelihood ratio test. A superior model was also expected to reduce inter-subject variability terms and/or residual error terms.

With the objective of increasing the stability of the model and reducing the uncertainty around the parameters of interest, the use of the Normal-Inverse Wishart Prior (NWPRI) approach was used in NONMEM¹⁹ to test the impact on the estimates of the fixed and random effects in the pharmacokinetic model under development. Primary PK parameters estimated with a previously developed model in adults²⁰ were used as prior information for the pharmacokinetic analysis of DFP in the target population.

Covariate analysis

Continuous and categorical covariates were tested during the analysis. The relationship between individual PK parameters (post-hoc or conditional estimates) and covariates was explored by graphical methods (plot of each covariate vs. each individual parameter). Relevant demographic covariates (body weight, height, age and gender) were entered one

by one into the population model (univariate analysis). After all significant covariates had been entered into the model (forward selection), each covariate was removed (backward elimination), one at a time. The model was run again and the objective function recorded. The likelihood ratio test was used to assess whether the difference in the objective function between the base model and the full (more complex) model was significant. The difference in -2Log likelihood (DOBJF) between the base and the full model is approximately χ^2 distributed, with degrees of freedom equal to the difference in number of parameters between the two hierarchical models. Because of the exploratory nature of this investigation, for univariate analyses, additional parameters leading to a decrease in the objective function of 3.84 was considered significant ($p < 0.05$). During the final steps of the model building, only the covariates which resulted in a difference of objective function of at least 7.88 ($p < 0.005$) were kept in the final model.

Model validation

The validation of the final pharmacokinetic model was based on graphical and statistical methods, including visual predictive checks¹⁷. Given the importance of the validation procedures for the subsequent use of a model for simulation purposes, in this study we have included a wide range of diagnostic methods to assess the accuracy of the parameter estimates and the predictive performance of the model¹⁸. Bootstrap was used to identify bias, stability and accuracy of the parameter estimates (standard errors and confidence intervals). The bootstrap procedures were performed in PsN v3.5.3 (University of Uppsala, Sweden)²¹, which automatically generates a series of new data sets by sampling individuals with replacement from the original data pool, fitting the model to each new data set. Subsequently, parameter estimates were used to simulate plasma concentrations in subjects with similar demographic characteristics, dosing regimens and sampling scheme as in the original clinical studies. Mirror plots were also generated to evaluate the variance-covariance structure of the parameters in the model, which is reflected by the degree of similarity between the original fit and the pattern obtained from the fitting of the simulated data sets using the final pharmacokinetic model.

PK bridging and dosing recommendations

To optimise the deferiprone dosing regimen in the target population, simulations were performed to achieve systemic exposure values similar to the adult reference population²⁰. Simulations were carried out to explore how differences in demographic covariates might affect steady-state exposure to deferiprone treatment. Sampling frequency and times were based on a serial sampling scheme for the purposes of estimating AUC, C_{max} and C_{ss} over the dosing interval. Integration of the concentration time data was applied according to the trapezoidal rule to ensure realistic estimates of variability. The adequacy of the simulated

dosing regimens was assessed graphically by determining the fraction of the paediatric population reaching systemic exposure comparable to the target value based on PKPD reference in adults.

A study duration of one week was chosen for the simulation. Each scenario consisted of 1000 simulations. Two dosing regimens were simulated in both populations: 75 and 100 mg/kg/day as three daily doses of 25 and 33.3 mg/kg respectively. The pharmacokinetic parameters of interest (AUC, C_{max} and C_{ss}) were measured after administration of the first dose on day 7.

A pharmacokinetic model developed in adult healthy volunteers²⁰ was used to simulate deferiprone exposure in the reference population. A population of 100 subjects (50 males and 50 females) with a body weight distribution of mean 55 and sd 7.5 kg was used to characterise a standard adult thalassaemic population.

The final PK model developed during this analysis was used to simulate deferiprone exposure in the population of interest. A population of 100 subjects (50 males and 50 females) with a body weight distribution of mean 16 and sd 2.0 kg was used to characterise a standard thalassaemic population of children below 6 years of age.

5.3 Results

Population Pharmacokinetic Modelling

Data from 18 evaluable children (9 males and 9 females) were used for the pharmacokinetic analysis. Patients were randomised to 3 dose levels (8.3, 16.7 and 33.3 mg/kg) with 6 patients assigned to each group. 16 patients were diagnosed with β -thalassaemia major and 2 with thalassodrepanocytosis. Mean (and sd) body weight, height and age of the children were respectively 16.08 (3.18) Kg, 98.95 (9.16) cm and 3.62 (1.33) years.

The pharmacokinetics of deferiprone after oral administration to paediatric patients was described by a one-compartment open model with first-order absorption and elimination processes. The absorption rate constant (K_a) represents a first order process. The disposition processes includes (apparent) clearance (CL/F) and (apparent) volume of distribution (V/F).

Between subject variability (BSV) was tested on each parameter, and was included in the final model on CL/F and V/F. An omega block was implemented in the estimation of BSV for CL/F and V/F, accounting for the expected correlation between these two parameters. The inclusion of the omega block significantly decreased the OBJF.

Different error models were tested to characterise residual variability; e.g., additive, proportional, exponential, combined, etc. The proportional error model provided the best results and was kept to describe the residual variability.

The use of the Normal-Inverse Wishart Prior (NWPRI) approach was used in NONMEM to estimate the fixed effect on the PK parameter K_a and the BSV for CL/F and V/F. The use of a prior allowed a better description of the data, reducing significantly the uncertainty around

the parameters above mentioned. The prior information was derived from a population PK analysis performed in healthy adults receiving deferiprone as a 100 mg/ml solution ²⁰. The following values were used for the different parameters: 8.2 h⁻¹ for Ka with an uncertainty of 4.02; 0.057 (23.87%) variation on CL/F and 0.0278 (16.67%) variation on V/F with an omega block of 0.0345. 54 degrees of freedom were chosen for the prior on the BSV parameters given that 55 individuals were used for the final population PK model in the healthy adults. During covariate model selection, after a visual explorative analysis of the correlations between covariates and model parameters, the effect of weight, height, gender, and age was tested on the different parameters. The inclusion of body weight on CL/F and V/F according to fixed allometric scaling ²² led to the highest improvement in the model fitting and allowed a better description of the data, increasing the model performance. The exponent was fixed to 0.75 and 1 for CL/F and V/F respectively. An overview of the final parameter estimates is provided in Table 1.

Table 1. Population pharmacokinetic parameters of deferiprone in children below 6 years of age and bootstrap results

Model predicted primary PK parameters				
	Estimate	SE	Bootstrap^a (mean)	CV (%)
CL/F (L/h)	8.3	0.569	8.30	8.07
V/F (L)	18.7	1.16	18.74	7.95
Ka (h-1)	9.13	1.41	8.91	10.54
WT on V/F Fix allom.	1 FIX	/	1 FIX	/
WT on CL/F Fix allom.	0.75 FIX	/	0.75 FIX	/
Error (prop)	0.0953	0.0182	0.0916	39.3
IIV CL/F^b	0.0644	0.0115	0.0642	11.37
IIV V/F^b	0.0392	0.0077	0.0393	13.23
Block CL-V	0.031	0.0058	0.0313	12.14
Model predicted secondary PK parameters stratified per dose level				
Median (5th and 95th quantiles)				
	Dose 1^c	Dose 2^d	Dose 3^e	
AUC₀₋₈ (µmol/L*h)	116.7 (90.6-129.0)	210.0 (173.1-266.6)	428.8 (291.4-547.8)	
Cmax (µmol/L)	61.7 (45.1-80.7)	119.8 (106.0-154.0)	229.5 (179.7-278.1)	
Tmax (h)	0.33 (0.19-0.92)	0.33 (0.21-0.63)	0.37 (0.27-0.42)	
Css (µmol/L)	2.1 (1.6-2.3)	3.7 (3.1-4.9)	7.7 (5.1-10.0)	
Cmin (µmol/L)	1.5 (0.92-2.6)	1.9 (0.79-5.5)	6.8 (3.1-13.9)	

^a 0 minimisation terminated out of 500; ^b Eta shrinkage was -11% and 0% for CL/F and V/F respectively; ^c 8.3 mg/kg; ^d 16.7 mg/kg; ^e 33.3 mg/kg

A bootstrap analysis was performed to assess model stability. The mean parameter estimates from the bootstrap analysis were found to be in close agreement with the final model estimates, and the CV values were found to be all below 15%, indicating that the final estimates are indeed reliable. Results of the bootstrap analysis can be found in Table 1. Internal model validation diagnostics were satisfactory. Individual predicted profiles and goodness-of-fit plots revealed that the model provides an adequate and non-biased description of the data, as shown in Figures 1 and 2.

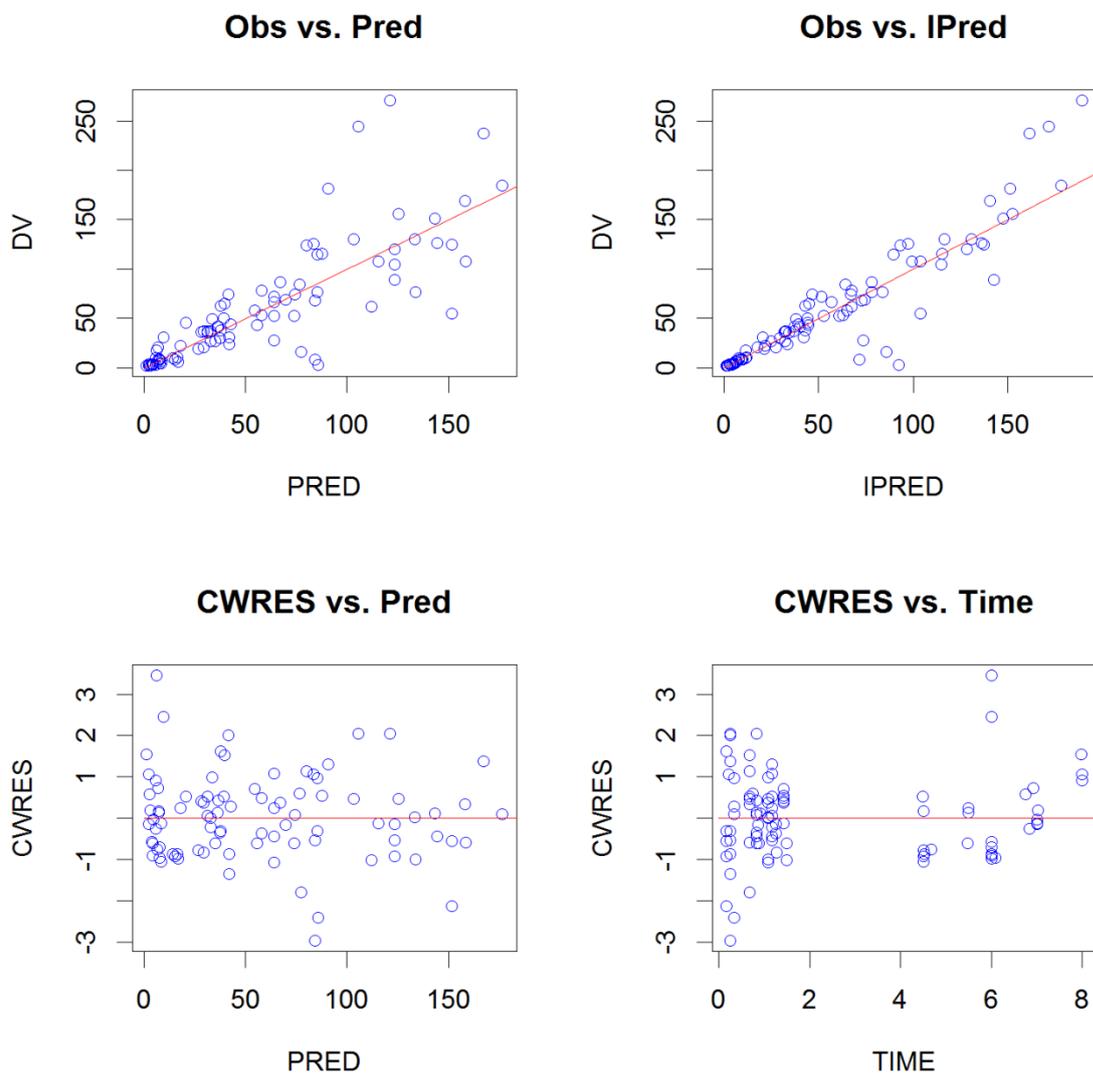
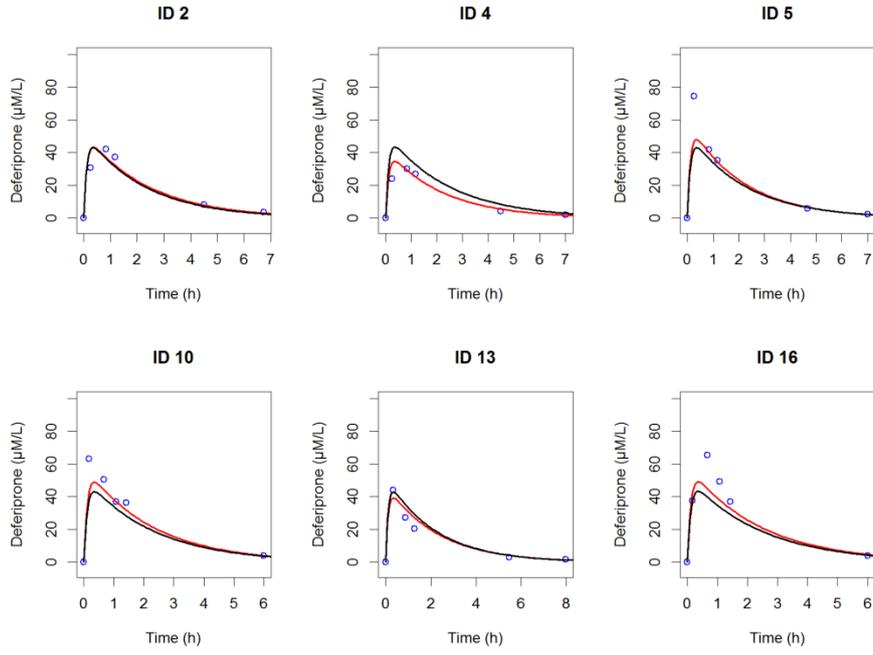


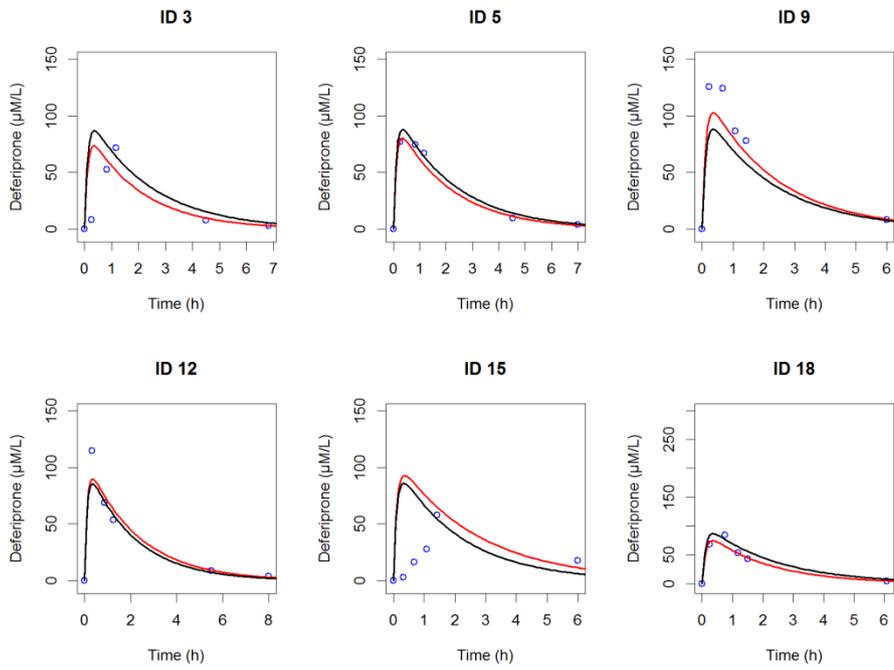
Figure 1. Goodness-of-fit plots. Upper panels show the observed data (Obs) vs. individual predictions (IPred) (left) and the observed data vs. population predictions (Pred) (right). Lower panels show the conditional weighted residuals (CWRES) vs. population predictions (left) and the CWRES vs time (left).

MODEL-BASED DOSING RECOMMENDATIONS OF DEFERIPRONE IN CHILDREN

A



B



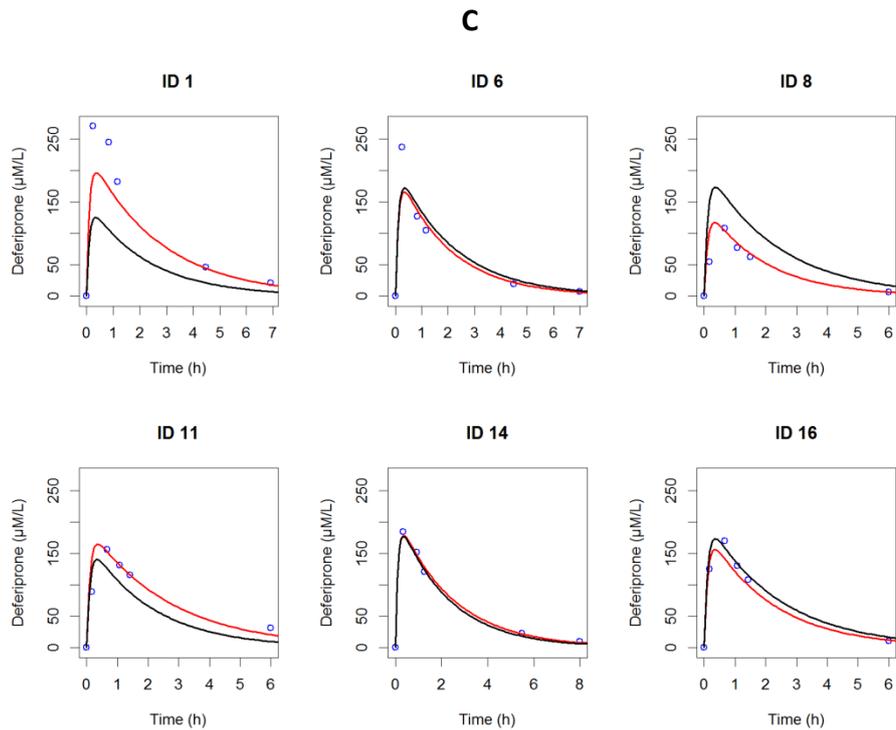


Figure 2. Individual plots: observed data are plotted using blue circles; the black solid line represents the population prediction (Pred) and the red solid line represents the individual predictions (IPred). Panel A shows patients in dose group 1 (8.3 mg/kg); panel B shows patients in dose group 2 (16.7 mg/kg); and panel C shows patients in dose group 3 (33.3 mg/kg).

In addition, NPDE summaries (Figure 3) show that the discrepancy between predicted and observed values can be assumed to be normally distributed. The predictive performance of the model in subsequent simulations was deemed critical to achieve the objective of our analysis. To this purpose, visual predictive checks were therefore used to assess whether the variance and covariance structures have been well characterised (Figure 4). Overall these diagnostic techniques confirm that the final model is suitable for the purposes of data simulation.

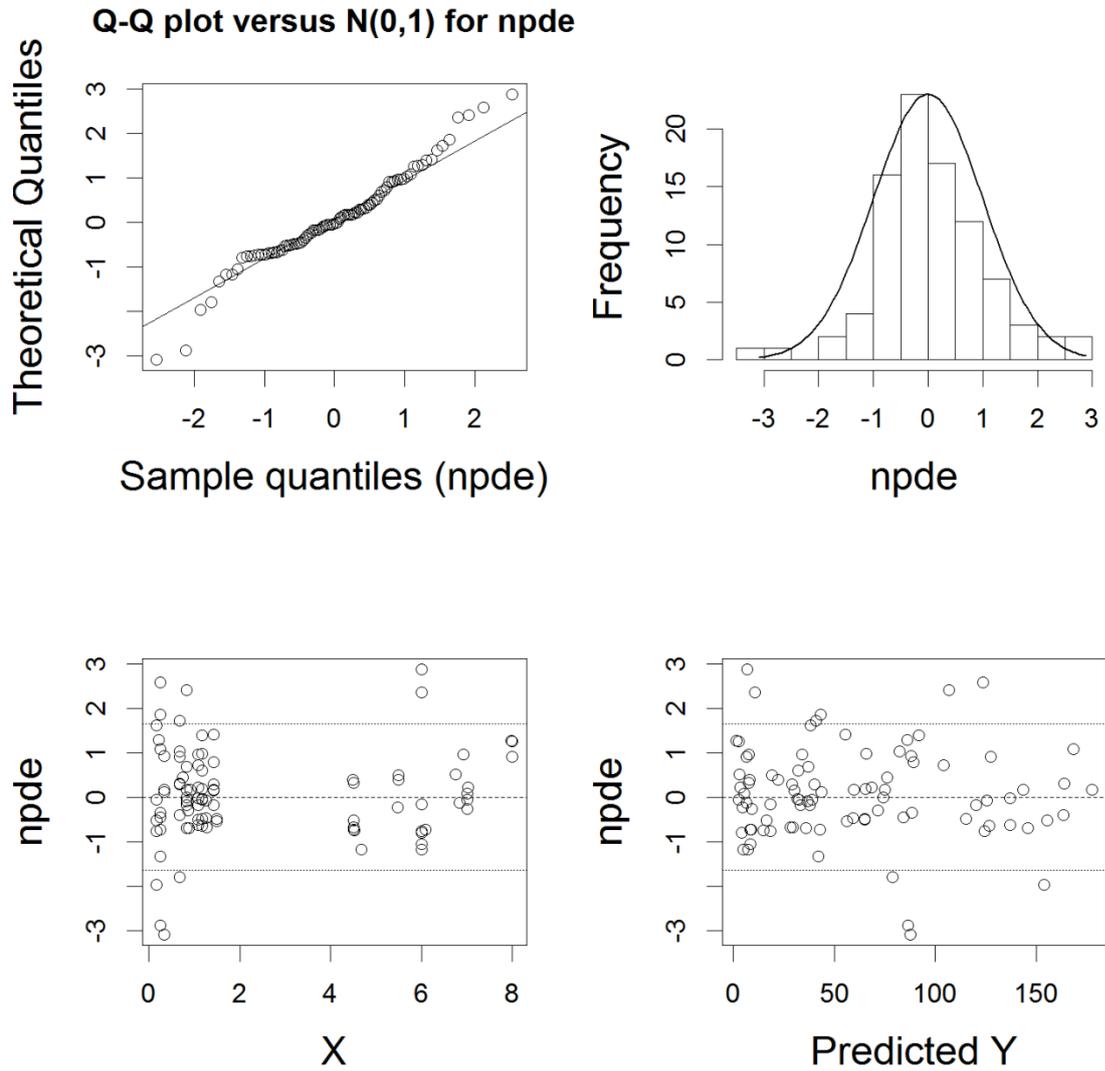


Figure 3. Normalised prediction distribution errors: upper panels show the QQ-plot of the distribution of the NPDEs for a theoretical $N(0, 1)$ distribution (left) and the histogram of the distribution of the NPDE together with the density of the standard normal distribution (right). Lower panels show the NPDEs vs. time (left) and NPDEs vs. individual predictions (right).

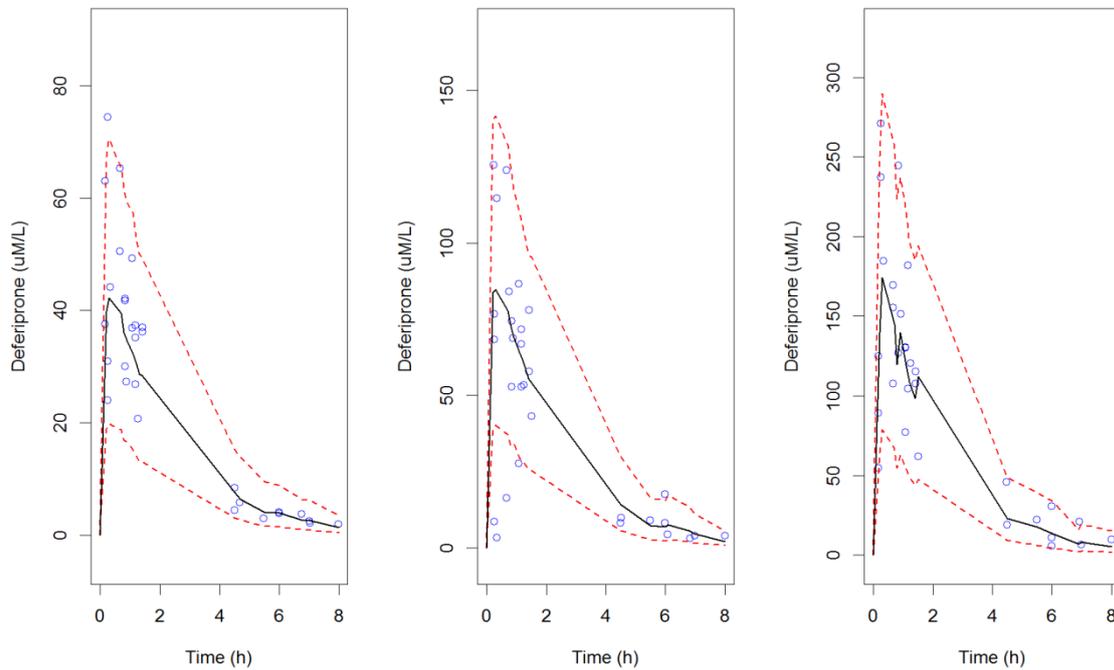


Figure 4. Visual Predictive Check (VPC): observed data are plotted using open circles; the black solid line represents the median of the simulated data; the dashed lines represent the 5th and 95th quantiles of the simulated data. The left, mid and right panels show respectively dose group 1 (8.3 mg/kg), 2 (16.7 mg/kg) and 3 (33.3 mg/kg).

PK bridging and dosing recommendations

The results of the simulations are shown in Figures 6 and 7 and Table 2. A similar exposure is achieved in adults and children in terms of AUC and C_{ss} when receiving the current recommended dosing regimen both at 75 and 100 mg/kg/day. The simulation generated a 29% increase in C_{max} in children when compared to the adult population.

The performance of an individualised dosing regimen was tested on the target population, but the results show that it does not change significantly the exposure in children when compared to the non-individualised one (at 75 mg/kg/day); not shown here.

Results suggest that the currently approved dosing regimen for the adult population is suitable also for children below 6 years of age in order to achieve a similar and effective exposure.

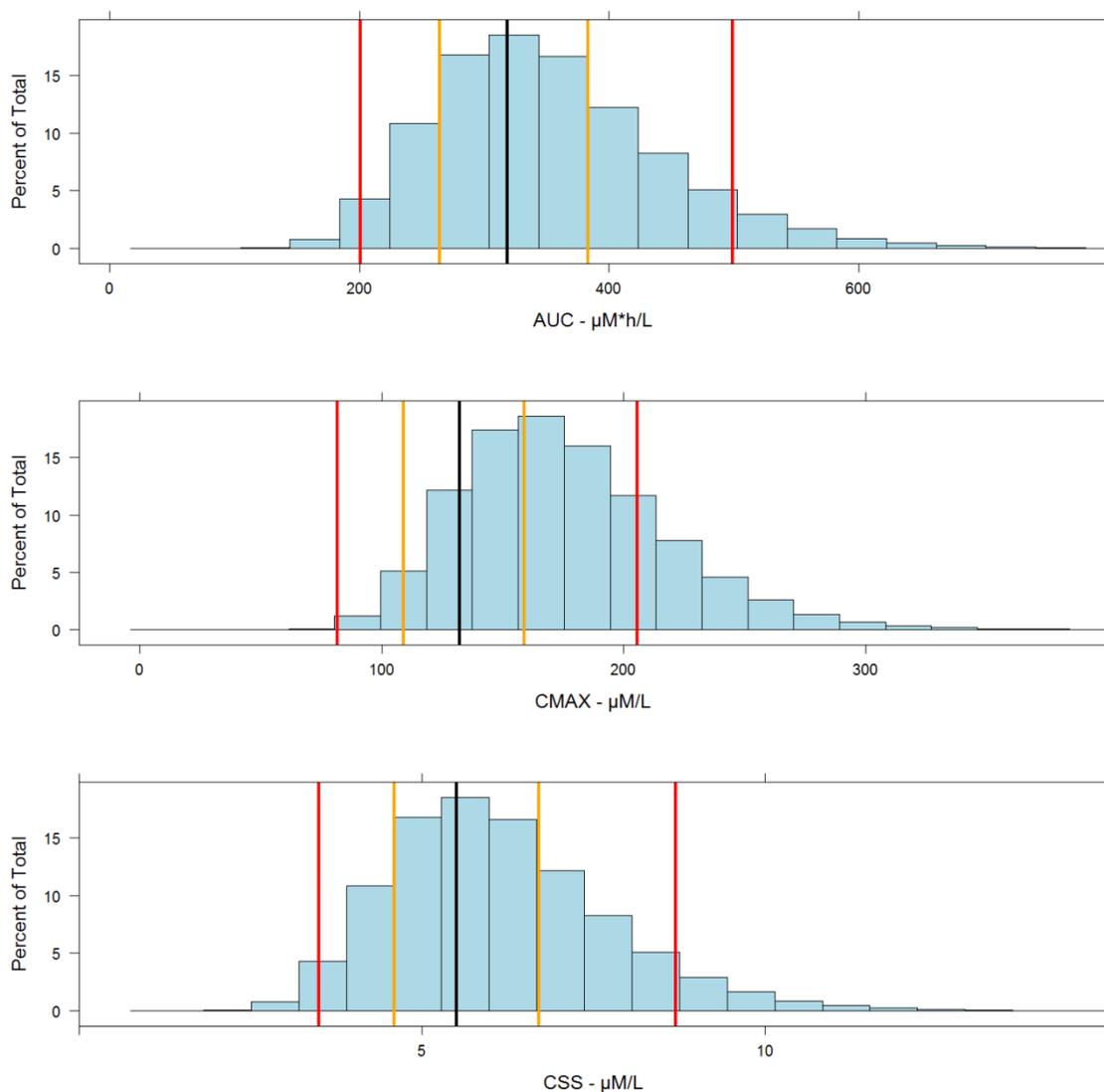


Figure 5. Predicted deferiprone exposure expressed as AUC 0-8 (upper panel), C_{max} (mid panel) and C_{ss} (lower panel) for children below 6 years of age receiving 75 mg/kg/day. The black line represents the median of the reference population (adult thalassaemic population), whereas the orange lines represent 1st and 3rd quartiles and the red lines represent 5th and 95th percentiles of the same reference population. Percent of total indicates the percentage of cases for each beam of 1000 simulations with 100 patients in each simulated trial.

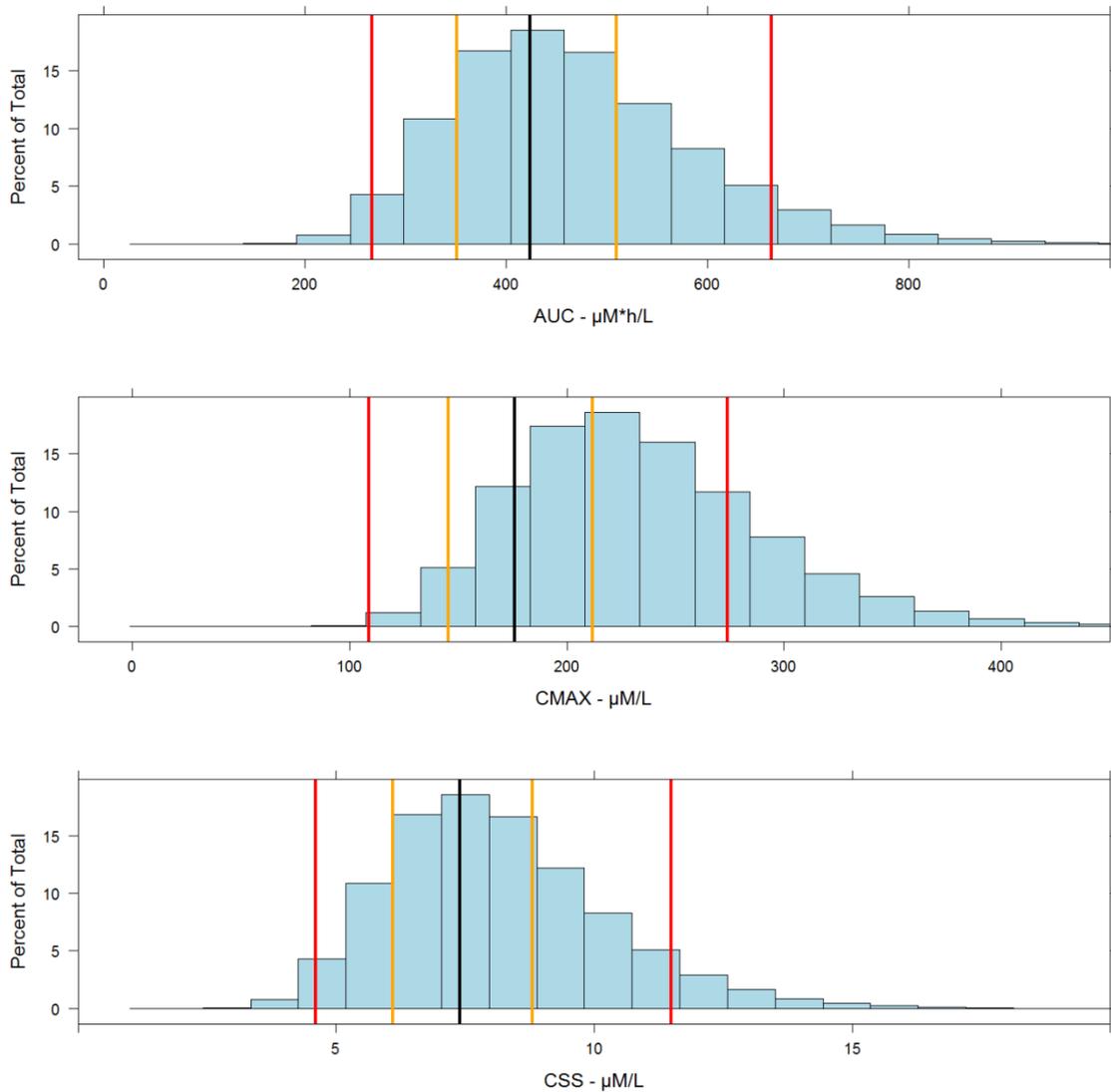


Figure 6. Predicted deferiprone exposure expressed as AUC 0-8 (upper panel), Cmax (mid panel) and C_{ss} (lower panel) for children below 6 years of age receiving 100 mg/kg/day. The black line represents the median of the reference population (adult thalassaemic population), whereas the orange lines represent 1st and 3rd quartiles and the red lines represent 5th and 95th percentiles of the same reference population. Percent of total indicates the percentage of cases for each beam of 1000 simulations with 100 patients in each simulated trial.

Table 2. Summary statistics of the simulation scenarios for the PK bridging study.

	75 mg/kg/day						100 mg/kg/day					
	Adults			Children			Adults			Children		
	AUC	Cmax	Css	AUC	Cmax	Css	AUC	Cmax	Css	AUC	Cmax	Css
Median	318.5	132.2	5.5	340.6	170.7	5.9	424.2	176.0	7.4	453.7	227.4	7.9
1st quartile	263.9	109.2	4.6	286.6	145.0	5.0	351.5	145.4	6.1	381.8	193.2	6.6
3rd quartile	383.0	159.0	6.7	404.7	200.5	7.0	510.0	211.9	8.8	539.0	267.1	9.4
5th quantile	200.4	81.6	3.5	223.2	114.9	3.9	266.9	108.7	4.6	297.3	153.1	5.2
95th quantile	499.0	205.6	8.7	520.0	253.0	9.0	664.0	273.9	11.5	693.0	337.0	12.0

AUC: $\mu\text{mol/L}\cdot\text{h}$; Cmax: $\mu\text{mol/L}$; C_{ss}: $\mu\text{mol/L}$

5.4 Discussion and Conclusion

Model-based approaches can be critical for therapeutic decisions when limited evidence is available. This is certainly the case for rare diseases such as haemoglobinopathies, especially when considering young paediatric patients, where practical and ethical constraints wisely imposed by regulatory authorities, make paediatric clinical investigation a true challenge^{23,24}. The lack of exhaustive experimental data available on the use of deferiprone in children including deferiprone pharmacokinetic data in children below 6 years of age hampered the ability to assess whether doses, used in adults, adjusted for weight, would produce comparable exposure in young children. The need for a better understanding of DFP behaviour in the paediatric population led to the establishment of the DEEP consortium (www.deep.cvbf.net). Within this project, a model-based approach has been used to overcome the specific challenge to better understanding DFP behaviour and allowing adequate dosage in the <6 years of age group, reducing at the same time the sampling burden on such a vulnerable population (i.e., by the use of optimal design techniques to increase the quality of the information gathered and by the use of population PK analysis in the presence of sparse sampling). Modelling and Simulations (M&S) techniques have become an invaluable tool for the evaluation of the dose rationale and personalisation of dosing regimens for subgroups of patients and special populations, allowing the characterisation and quantification of the contribution of different sources of variability to an agent's overall pharmacokinetic properties. Furthermore, continuous emphasis has been placed on the need for evidence-based clinical and regulatory decisions, where modelling and simulation is becoming more and more an essential component²⁵⁻²⁷.

Pharmacokinetic modelling

The pharmacokinetics of deferiprone after oral administration to paediatric patients was successfully characterised by a model-based approach. As shown in the results section a one-compartment open model with first-order absorption and elimination processes

described satisfactorily the PK profile of the drug under investigation, allowing precise and accurate characterisation of the main PK parameters of interest (Table 1). Body weight was found to be a significant predictor of changes in the distribution and elimination processes of the drug; the relationship with CL/F and V/F was described by fixed allometric scaling. Furthermore, the use of prior information in the adult population allowed a more stable characterisation of the absorption profile, showing once more how M&S can overcome the limited evidence generated in the clinical study.

Dosing recommendations

Bridging concepts are applied in this context to evaluate the exposure in the paediatric population as compared to efficacious exposure in adults. Using the model developed for this population, and a previously developed model in the adult population²⁰, simulations were performed to compare PK exposure between children below 6 years of age and a standard thalassaemic adult population. As shown in Figures 6 and 7, AUC and C_{ss} distributions are comparable at 75 mg/kg/day and 100 mg/kg/day respectively, whereas an increase in peak concentrations (C_{max}) is predicted in children. This increase is most probably due to differences in the volume of distribution between the two groups, and is expected to have limited clinical implications. Overall exposure (AUC and C_{ss}) is the determinant of the response, and changes in C_{max} are not expected to modify the safety profile of the drug. This is confirmed in literature where previous studies in children exposed to a 100 mg/kg/day dosing regimen have safety profiles similar to those reported in adults^{28–30}.

In conclusion, based on these findings, a dosing regimen of 25 mg/kg t.i.d. (75 mg/kg/day) is recommended for children aged from 1 month to < 6 years, with the possibility of titration up to 33.3 mg/kg t.i.d. (100 mg/kg/day), if necessary. Noticeable, this dosage will be used to conduct further efficacy-safety comparative phase III study and will be also adopted in any SmPC possible modifications.

References

1. R.Gibbons, D.R.Higgs, J.M.Old, Olivieri Nancy F., Thein Swee Lay WGW. The Thalassaemia Syndromes - Fourth Edition. *Blackwell Sci.* 2001;
2. Cunningham MJ, Macklin E a, Neufeld EJ, Cohen AR. Complications of beta-thalassemia major in North America. *Blood.* 2004;104(1):34–9.
3. Borgna-Pignatti C, Cappellini MD, De Stefano P, et al. Survival and complications in thalassemia. *Ann. N. Y. Acad. Sci.* 2005;1054:40–7.
4. Borgna-Pignatti C, Cappellini MD, De Stefano P, et al. Cardiac morbidity and mortality in deferoxamine- or deferiprone-treated patients with thalassemia major. *Blood.* 2006;107(9):3733–7.
5. Galanello R, Campus S. Deferiprone chelation therapy for thalassemia major. *Acta Haematol.* 2009;122(2-3):155–64.
6. Barman Balfour J a, Foster RH. Deferiprone: a review of its clinical potential in iron overload in beta-thalassaemia major and other transfusion-dependent diseases. *Drugs.* 1999;58(3):553–78.
7. Fassos FF, Klein J, Fernandes D, Matsui D, Olivieri NF KG. The pharmacokinetics and pharmacodynamics of the oral iron chelator deferiprone (L1) in relation to hemoglobin levels. *Int J Clin Pharmacol Ther.* 1996;34(7):288–92.
8. EMA. Deferiprone Summary of Product Characteristics.
9. Benoit-Biancamano MO, Connelly J, Villeneuve L, Caron P, Guillemette C. Deferiprone Glucuronidation by Human Tissues and Recombinant UDP Glucuronosyltransferase 1A6 : An in Vitro Investigation of Genetic and Splice Variants. *Drug metab dispos.* 2009;37(2):322–329.
10. Diav-citrin O. Oral iron chelation with deferiprone. *NEW Front. Pediatr. DRUG Ther.* 1997;44(1):235–247.
11. Hoffbrand AV, Cohen A, Hershko C. Role of deferiprone in chelation therapy for transfusional iron overload. *Blood.* 2003;102(1):17–24.
12. Victor Hoffbrand a. Deferiprone therapy for transfusional iron overload. *Best Pract. Res. Clin. Haematol.* 2005;18(2):299–317.
13. Limenta LMG, Jirasomprasert T, Tankaniltert J, et al. UGT1A6 genotype-related pharmacokinetics of deferiprone (L1) in healthy volunteers. *Br. J. Clin. Pharmacol.* 2008;65(6):908–16.
14. Stobie S, Tyberg J, Matsui D, Fernandes D, Klein J, Olivieri N, Bentur Y KG. Comparison of the pharmacokinetics of 1,2-dimethyl-3-hydroxypyrid-4-one (L1) in healthy volunteers, with and

- without co-administration of ferrous sulfate, to thalassemia patients. *Int J Clin Pharmacol Ther Toxicol.* 1993;31(12):602–5.
15. al-Refaie FN, Sheppard LN, Nortey P, Wonke B, Hoffbrand a V. Pharmacokinetics of the oral iron chelator deferiprone (L1) in patients with iron overload. *Br. J. Haematol.* 1995;89(2):403–8.
 16. Matsui D. Relationship between the pharmacokinetics and iron excretion pharmacodynamics of the new oral iron chelator 1,2-dimethyl-3-hydroxypyrid-4-one in patients with thalassemia. *Clin. Pharmacol. Ther.* 1991;50(3):294–8.
 17. Hooker AC, Staats CE, Karlsson MO. Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method. *Pharm. Res.* 2007;24(12):2187–97.
 18. Comets E, Brendel K, Mentré F. Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. *Comput. Methods Programs Biomed.* 2008;90(2):154–66.
 19. Boeckmann AJ, Sheiner LB, Beal SL. NONMEM Users Guide - Part VIII. 2011.
 20. Bellanti F, Danhof M, Della Pasqua O. Population pharmacokinetics of deferiprone in healthy subjects. *Br. J. Clin. Pharmacol.* 2014;78(6):1397–406.
 21. Lindbom L, Ribbing J, Jonsson EN. Perl-speaks-NONMEM (PsN)--a Perl module for NONMEM related programming. *Comput. Methods Programs Biomed.* 2004;75(2):85–94.
 22. Anderson BJ, Allegaert K, Holford NHG. Population clinical pharmacology of children: general principles. *Eur. J. Pediatr.* 2006;165(11):741–6.
 23. EMA. Paediatric Workshops - http://www.emea.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000416.jsp&mid=WC0b01ac0580118a19.
 24. ICH Topic E11. http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002926.pdf. 2001;
 25. Bellanti F, Della Pasqua O. Modelling and simulation as research tools in paediatric drug development. *Eur. J. Clin. Pharmacol.* 2011;67 Suppl 1:75–86.
 26. Manolis E, Osman TE, Herold R, et al. Role of modeling and simulation in pediatric investigation plans. *Paediatr. Anaesth.* 2011;21(3):214–21.
 27. 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.

28. Ceci A, Baiardi P, Felisi M, et al. The safety and effectiveness of deferiprone in a large-scale, 3-year study in Italian patients. *Br. J. Haematol.* 2002;118(1):330–6.
29. ElAlfy MS, El Alfy M, Sari TT, et al. The safety, tolerability, and efficacy of a liquid formulation of deferiprone in young children with transfusional iron overload. *J. Pediatr. Hematol. Oncol.* 2010;32(8):601–5.
30. Makis A, Chaliasos N, Alfantaki S, Karagouni P, Siamopoulou A. Chelation therapy with oral solution of deferiprone in transfusional iron-overloaded children with hemoglobinopathies. *Anemia.* 2013;2013:121762.

SECTION III

Integrated evaluation of efficacy and safety
by modelling and simulation

CHAPTER 6

A disease model for iron overload in patients affected by transfusion-dependent diseases

F Bellanti, M Muliaditan, M Danhof, and O Della Pasqua

Submitted for publication

Summary

The understanding of iron overload dynamics and its progression is essential to establish an adequate therapeutic intervention in patients affected by transfusion-dependent diseases. The main objective of this analysis is to develop a disease model for iron overload on the basis of available literature data. A thorough literature search was performed in Pubmed to retrieve all pertinent publications that would allow characterising the different aspects of the disease. At first, the turnover of serum ferritin in healthy individuals was described by an indirect response model. Subsequently, the effect of blood transfusions on serum ferritin levels was quantified according to an Emax model that depicts the non-linearity of the relationship. Finally, the relationship has been integrated as an additive conversion rate in the turnover model to account for disease progression. Internal model validation diagnostics were satisfactory and visual predictive checks reveal that the model provides an adequate and non biased description of the data. In conclusion, a disease model for iron overload was successfully developed. The relationship between blood transfusions and serum ferritin levels was quantified for the first time through a model-based approach. This model puts the basis for a more structured evaluation of therapeutic intervention in this patient population.

6.1 Introduction

Patients affected by transfusion-dependent diseases, such as beta-thalassaemia major or sickle cell disease, require regular red blood cell (RBC) transfusions to survive ¹⁻⁷. Without the chronic transfusion regimen, patients would die before the third decade of life. Based on the Guidelines for the Clinical Management of Thalassaemia ⁷, transfusions should aim at maintaining a pre-transfusion haemoglobin (Hb) level between 9 and 10 g/dl and a post-transfusion level of 14 to 15 g/dl. The most common transfusion interval in these patients is once every two to four weeks (equal to two to three blood units per three weeks) ^{2,4,5,7,8}.

Iron overload in patients affected by transfusion-dependent diseases

As generally known, iron is recycled within the body and the body itself does not have the capacity to remove the excess of iron that is introduced from continuous blood transfusions. In normal conditions (Figure 1), iron entry into the cells is regulated by the uptake of iron-transport protein transferrin from the plasma. Chronic blood transfusions induce an increased iron exposure from macrophages, resulting into saturation of transferrin transport capacity. This leads to the release of Non-Transferrin Bound Iron (NTBI) in the plasma which can enter important cells (e.g., heart and liver cells) resulting over time into tissue accumulation. Iron is stored in tissues mainly into ferritin complexes. Once ferritin storage capacity is overwhelmed, small clusters of ferritin particles are formed and are degraded by the lysosomes leading to the formation of insoluble masses of hemosiderin ⁹⁻¹⁵. Over time this accumulation would cause severe organ damage ¹⁶⁻²¹.

Even though a significant improvement has been achieved in the management of the chronic transfusion regimen in the past decades, the therapy will eventually lead to a series of complications. Iron overload is the most common and relevant one and it is associated with several (lethal) co-morbidities such as cardiac dysfunction, liver fibrosis, hypogonadism, hypothyroidism, hypoparathyroidism and diabetes mellitus ^{18,20}. Cardiac disease caused by myocardial siderosis is the most relevant complication, causing death in 71% of the patients affected by transfusion-dependent diseases ¹⁷.

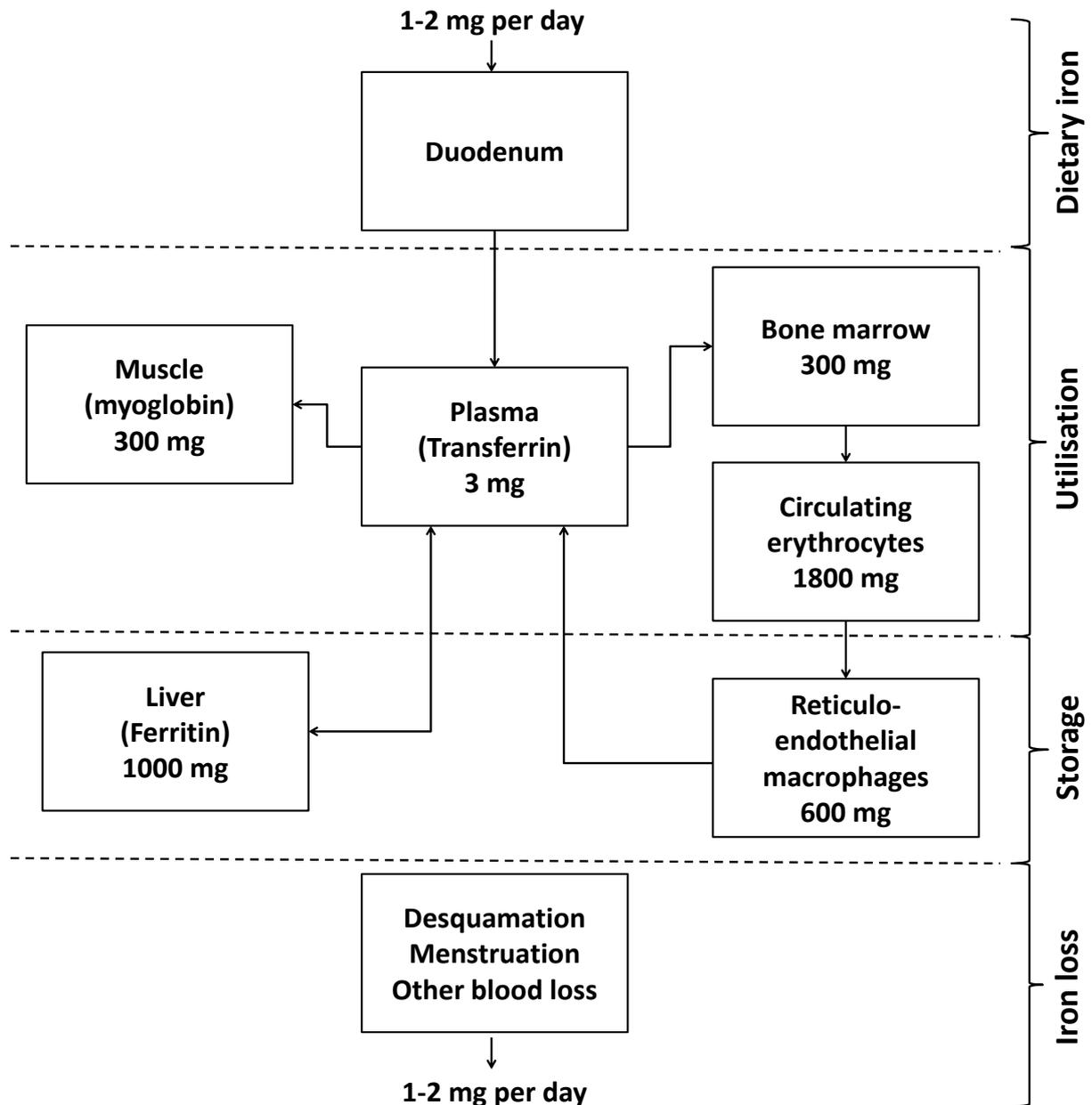


Figure 1. Iron homeostasis. Adapted from Andrews et al.¹⁶.

In the absence of an innate mechanism that allows removing iron excess from the body, treatment with iron chelators is therefore essential to prevent iron accumulation and related complications²²⁻²⁵. Iron chelators possess overall a similar mechanism of action. They act by 1) preventing the uptake of NTBI into organs such as liver and heart; 2) chelating intracellular iron and thus preventing its incorporation into ferritin; or 3) intercepting iron released from degraded ferritin²⁶.

Clinical assessment of iron overload

There are several clinical measures to evaluate the disease state of iron overload. The most common is the biomarker serum ferritin, due to its strong correlation with total body iron stores²⁷. As a single clinical endpoint however, serum ferritin is not always reliable. It could also be influenced by other factors such as inflammatory disorders and liver disease²⁸. On the other hand, serial measurements of serum ferritin are still the easiest and least invasive method to evaluate iron overload and efficacy of chelation therapy.

Other assessment methods for iron status focus more on tissue specific accumulation. Liver iron concentration is considered as the gold standard for the evaluation of iron overload due to a high correlation with total body iron accumulation²⁹. However, determination of liver iron concentration requires an invasive technique with complications and risks of false negative results³⁰. Magnetic Bio-Susceptometry (SQUID) is another option for measurement of liver iron accumulation³¹. However, it is only available in a limited numbers of centres worldwide. Furthermore, cardiac complications due to iron accumulation in the heart have been associated with 50-70% of deaths in thalassaemia major patients, mainly at young age³². Methods that were developed for cardiac monitoring were based on keeping serum ferritin and LIC level below a certain threshold (<2500 µg/L and <7 mg/kg dwt respectively) that was associated with decreased cardiac risks. However, this method proved not to be sufficient for effective intervention, since any dysfunctions were often identified at relative late stage. In recent years, Magnetic Resonance Imaging (MRI) techniques for assessing iron loading in the liver and heart have been introduced and validated for the evaluation of tissue specific accumulation³³.

Iron overload is thus a rather complex process, and the understanding of the dynamics of the disease and its progression is essential for an adequate improvement of the therapeutic intervention. Several clinical questions are still not fully understood, e.g. how much time is required in order to observe a true response in the patient, or in order to reach clinically acceptable serum ferritin levels (i.e. about 2500 µg/L). As generally recognised, ferritin reflects what happens at the organ level only up to a certain threshold. Above this threshold other mechanisms intervene (inflammatory disorders, liver status)^{27,28} that influence the relationship between serum ferritin and body iron accumulation and the iron interchange between organs and the circulatory system. This project puts its main focus on the use of model-based approach to gain more insights in key factors that play a role in iron overload. The specific objective is to develop a disease model on the basis of available literature data. In particular we aim at quantifying the impact of blood transfusions on the changes in serum ferritin levels.

6.2 Methods

Data

A thorough literature search was performed in Pubmed to retrieve all pertinent publications that would allow the development of the disease model. Stepwise mining and pooling of published data was subsequently performed to characterise the different aspects of iron overload. Data published by Dawkins et al.³⁴ were used to quantify the turnover of serum ferritin in healthy individuals; whereas clinical data published by Worwood et al., George et al. and Letsky et al.³⁵ were pooled to evaluate the impact of blood transfusions on serum ferritin levels in untreated patients (i.e., patients not receiving chelation therapy).

Modelling

The software R (v.2.14.0) was used for statistical summaries and literature data extraction (complemented with R Digitize Package³⁶), as well as data manipulation and preparation for modelling purposes. Nonlinear mixed effects modelling was performed in NONMEM version 7.2 (Icon Development Solutions, USA).

Model building criteria included: (i) successful minimisation, (ii) standard error of estimates, (iii) number of significant digits, (iv) termination of the covariance step, (v) correlation between model parameters and (vi) acceptable gradients at the last iteration. Comparison of hierarchical models was based on the likelihood ratio test. Goodness of fit was assessed by graphical methods, including population and individual predicted vs. observed concentrations, conditional weighted residual vs. observed concentrations and time, correlation matrix for fixed vs. random effects, correlation matrix between parameters and covariates and normalised predictive distribution error (NPDE)³⁷.

The validation of the final model was based on graphical and statistical methods, including visual predictive checks³⁸. Bootstrap was used to identify bias, stability and accuracy of the parameter estimates (standard errors and confidence intervals). The bootstrap procedures were performed in PsN v3.5.3 (University of Uppsala, Sweden)³⁹, which automatically generates a series of new data sets by sampling individuals with replacement from the original data pool, fitting the model to each new data set.

Iron homeostasis in healthy individuals (basal ferritin turnover)

To quantify serum ferritin changes in healthy individuals, data from 14 subjects were extracted from literature³⁴ and combined into a single dataset. Data are presented in Table 1 and Figure 2. A turnover model was tested to describe serum ferritin profiles in this population:

$$\frac{dFERRITIN}{dt} = Kin - Kout \times FERRITIN$$

where K_{in} is the basal zero-order production rate of ferritin and K_{out} is the basal first-order degradation rate of ferritin.

Table 1. Summary of serum ferritin levels in healthy individuals: data is presented by study duration.

Study duration	Serum Ferritin ($\mu\text{g/L}$)	
	Mean \pm s.d	Range ($\mu\text{g/L}$)
7 weeks (N=9)	49.67 \pm 25.95	8.53 – 97.60
24 hours (N=5)	67.71 \pm 31.62	19.7 – 119.4

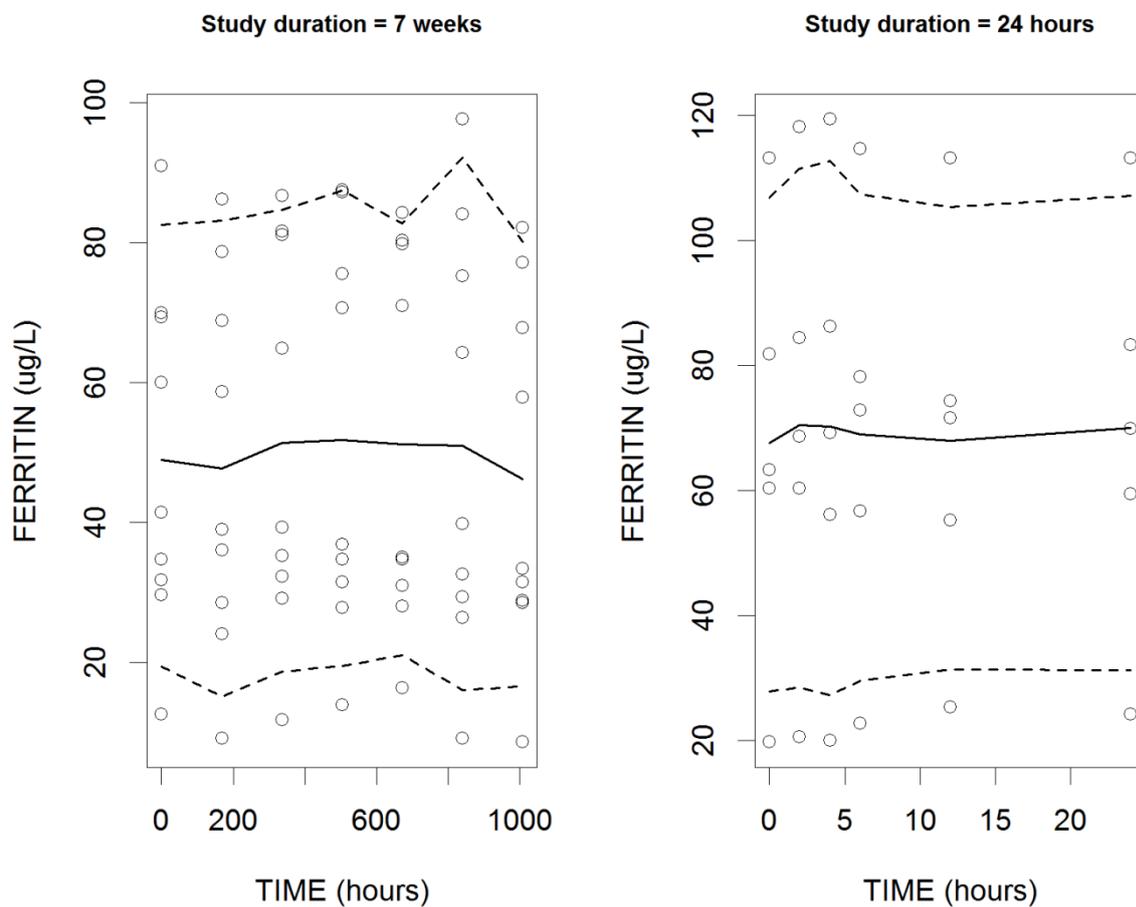


Figure 2. Serum ferritin changes over time in healthy individuals. Individual profiles in 14 healthy individuals presented as mean (solid line) and 5th and 95th percentiles (dashed lines). Left panel: serum ferritin profiles during an observational period of 7 weeks (N=9). Right panel: serum ferritin profiles during an observational period of 24 hours (N=5).

Relationship between serum ferritin and cumulative blood units (effect of transfusions)

Data containing serum ferritin levels in untreated patients were extracted and pooled from literature^{35,40,41} into a single dataset (Figure 3, right panel). Relevant information regarding the study population mentioned in the published articles is summarized in Table 2.

Table 2. Summary of serum ferritin levels and transfusion history in patients affected by transfusion-dependent diseases not receiving iron chelation therapy.

Study	N	Serum ferritin ($\mu\text{g/L}$)		Transfusion history (cum. blood units)	
		Mean \pm s.d	Range	Mean \pm s.d	Range
Overall	188	4271.55 \pm 3003.76	353-18780	153.8 \pm 106.42	4-502
Worwood et al.	116	5023.4 \pm 2512.38	445-14120	193.1 \pm 107.4	4-502
Letsky et al.	24	4902.63 \pm 4603.18	447.4-18780	116.5 \pm 96.3	4-278
George et al.	48	2694.64 \pm 2694.64	353-9046	77.6 \pm 43.05	13-224

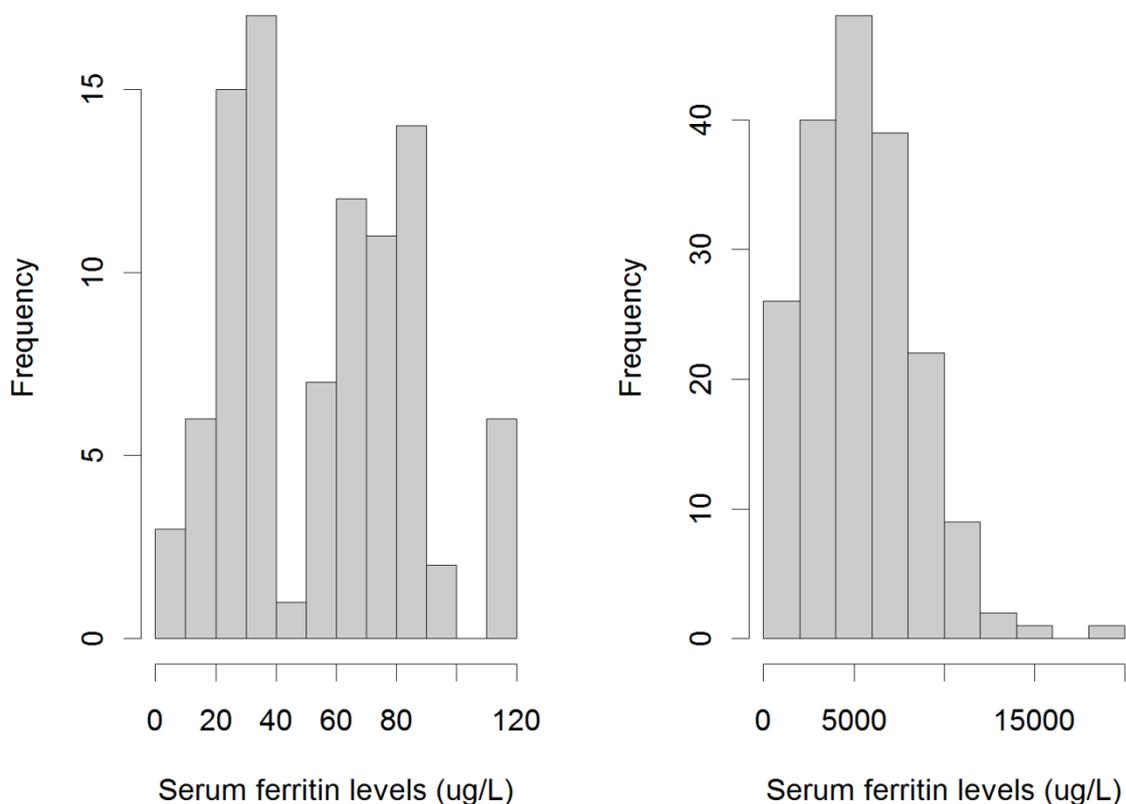


Figure 3. Comparison of the distributions of serum ferritin levels in healthy individuals (left panel) and patients affected by transfusion-dependent diseases not receiving iron chelation therapy (right panel).

Information regarding the volume of blood per unit was only available for the work carried out by Worwood et al. (500 ml per unit) and by George et al. (350 ml per unit). To ensure that equal volume of blood per transfusion was taken into account for the entire cohort, cumulative amount of blood units was normalized to a volume of 500 ml per blood unit.

Given the non-linear nature of the relationship between serum ferritin and cumulative blood units, an Emax model was tested to describe the relationship as described below:

$$FERRITIN = E_0 + \frac{Emax \times BU}{(Ebu_{50} + BU)}$$

where, E_0 represents baseline serum ferritin levels when no transfusion has yet occurred, $Emax$ the maximum serum ferritin levels at saturation and Ebu_{50} the cumulative amount of blood units (BU) when 50% of saturation is reached.

Ferritin data were log transformed for the analysis and the whole dataset was randomly divided into two subsets, resulting into 2/3 of the data used for the model building and 1/3 of the data preserved for external validation.

Integration of the effect of blood transfusions in the turnover model: disease model for iron overload

Once the relationship between cumulative blood units and serum ferritin was quantified, it was our goal to integrate this information within the turnover model. Our intent was to translate the relationship into a rate that would affect the basal ferritin production rate. Given that information on time was not provided in the data used to quantify the relationship, and given that we were mainly interested in translating the population profile, we assumed a constant interval of three weeks between subsequent units of blood transfused. This is on average the case in patients affected by transfusion-dependent diseases^{2,4,5,7,8}.

Assuming this constant time interval we performed a simulation-estimation analysis to quantify the impact of blood transfusions on the production rate of serum ferritin. The simulations were performed using the Emax model in the range of 5 to 450 cumulative blood units and subsequently the simulated data were fitted with the turnover model where basal K_{in} and K_{out} were fixed and a new production rate (CRT = conversion rate) was estimated (Figure 4). This rate was non-linearly correlated to actual ferritin levels according to the following equation:

$$CRT = SCL \times e^{-SHP \times FERRITIN}$$

where SCL is a scaling factor and SHP is the shape factor of the correlation. The conversion rate was integrated in an additive manner in the turnover model as follows:

$$\frac{dFERRITIN}{dt} = K_{in} + \mathbf{CRT} - K_{out} \times FERRITIN$$

CRT represents the effect of the disease (chronic transfusion therapy) on the biomarker ferritin.

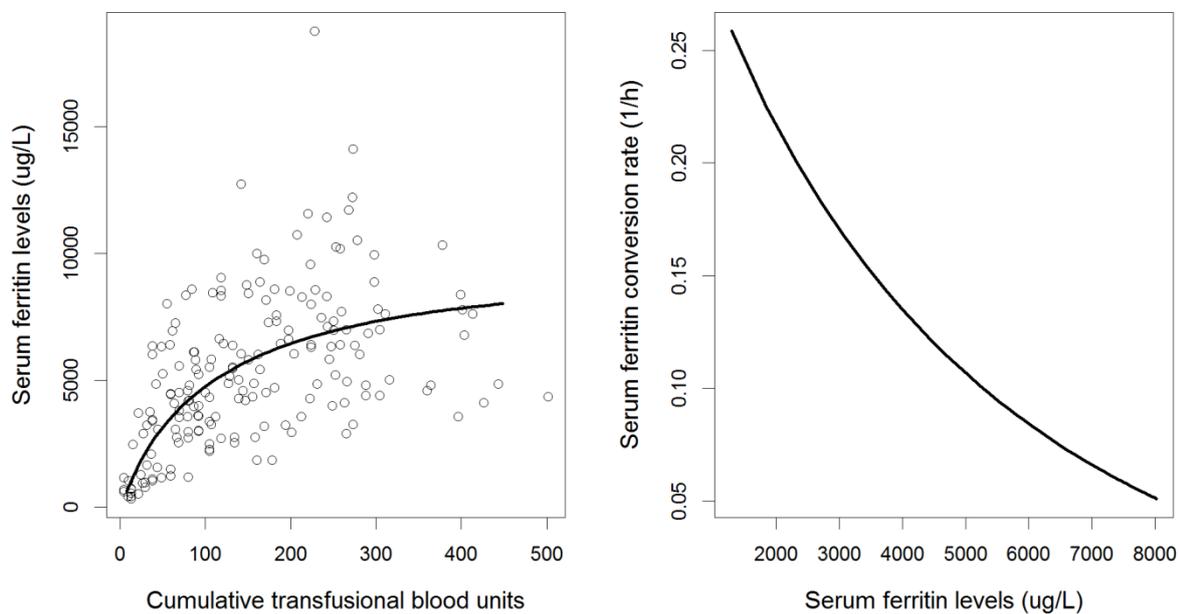


Figure 4. Stepwise integration of the effect of blood transfusions on serum ferritin levels in the turnover model developed on healthy subjects. Left panel: open circles represent the observed data and solid line represents the fitting of the relationship between cumulative blood units and serum ferritin levels. Right panel: negative relationship between serum ferritin conversion rate in the presence of chronic transfusion regimen and serum ferritin levels. The relationship has been derived from the one presented on the left panel assuming a constant time interval between consecutive blood units transfused.

6.3 Results

Iron homeostasis in healthy individuals (basal ferritin turnover)

An indirect response model was developed to describe basal serum ferritin turnover in healthy individuals. The zero-order production rate constant (K_{in}) and the first-order degradation rate constant (K_{out}) were successfully estimated; inclusion of inter-individual variability on K_{in} allowed a better description of the data. Parameter estimates and

bootstrap results are shown in Table 3. Goodness-of-Fit (GoF) (Figure 7, see Appendix) plots as well as Normalized Prediction Distribution Errors (NPDE) (Figure 8, see Appendix) confirm the suitability of the model in describing adequately the data.

Table 3. Final model parameters for turnover model of serum ferritin in healthy individuals.

Parameters	Estimates	Bootstrap (mean)	CV (%)
Kin ($\mu\text{g}/\text{day}$)	0.00625	0.00416	34.2
Kout (day^{-1})	0.000137	0.0000875	32.7
Eta on Kin	0.367	0.329	40.8
Residual error	0.000658	0.000641	31.2

Relationship between serum ferritin and cumulative blood units (effect of transfusions)

An Emax model was used to describe the relationship between serum ferritin levels and cumulative blood units. Inter-individual variability was estimated on the E_{bu50} parameters. The model allowed accurately quantifying the relationship; final parameter estimates are provided in Table 4 together with the estimates obtained with the external set of data. Goodness of Fit (Figures 9 and 11, see Appendix) plots as well as NPDE (Figures 10 and 12, see Appendix) reveal that the model provides a suitable description of the data.

Table 4. Summary of estimated relationship between cumulative blood units and serum ferritin levels in patients affected by transfusion-dependent diseases not receiving iron chelation therapy.

Parameters	Estimates	External Validation
E_0 ($\mu\text{g}/\text{L}$)	5.81	5.62
E_{max} ($\mu\text{g}/\text{L}$)	9.17	8.88
E_{bu50}	26.5	16
Eta on E_{bu50}	0.554	0.63
Residual error	0.0075	0.15

Integration of the effect of blood transfusions in the turnover model: disease model for iron overload

At first, ferritin levels corresponding to a range of cumulative blood units of 5 to 450 were simulated with the Emax model previously developed. Secondly, the data were fitted using an integrated model that consisted of the turnover model where basal Kin and Kout were fixed to the values estimated in the healthy population and an additive ferritin production rate (CRT) that was estimated during this process. The conversion rate (CRT) was non-linearly correlated to actual ferritin concentration. Final parameter estimates are provided in Table 5, whereas a schematic representation of the model is shown in Figure 5.

Table 5. Summary of final parameter estimates for the disease model of iron overload: integration of the effect of blood transfusions on serum ferritin levels in the turnover model developed on healthy subjects.

Parameters	Estimates
K_{in} ($\mu\text{g}/\text{h}$)	0.000208 FIX
K_{out} (h^{-1})	0.00000458 FIX
SHP (h^{-1})	0.00026
SCL ($\mu\text{g}/\text{h}$)	0.383
Residual error	0.0014

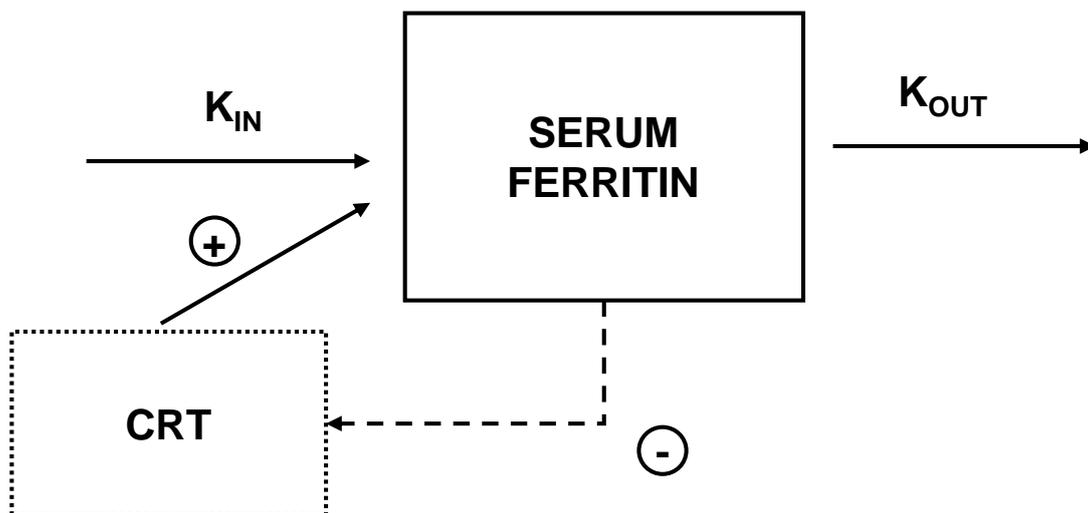


Figure 5. Disease model for iron overload. K_{in} and K_{out} represent respectively the basal zero order production rate and first order degradation rate of ferritin in healthy individuals. CRT represents the serum ferritin conversion rate in patients undergoing chronic transfusion therapy, which reflects the impact of the disease (blood transfusions) on serum ferritin levels. The dashed line represents the negative feedback that serum ferritin has on CRT.

Given the nature of the simulation, it was only possible to quantify the mean population profile for the integrated model. Inter-individual variability was added in a systematic manner to evaluate whether the model would capture the variation in the original data. Visual predictive checks (with the inclusion of 50% variability on both SCL and SHP) show that the model allows describing the data in an adequate and not biased manner (Figure 6

left panel). In addition, Figure 6 (right panel) shows simulated profiles of serum ferritin over a period of 10 years in a virtual patient not receiving iron chelation therapy. The simulations provide insights on how the impact of the disease changes when patients start at different baseline levels, and allow quantifying the true underlying disease progression.

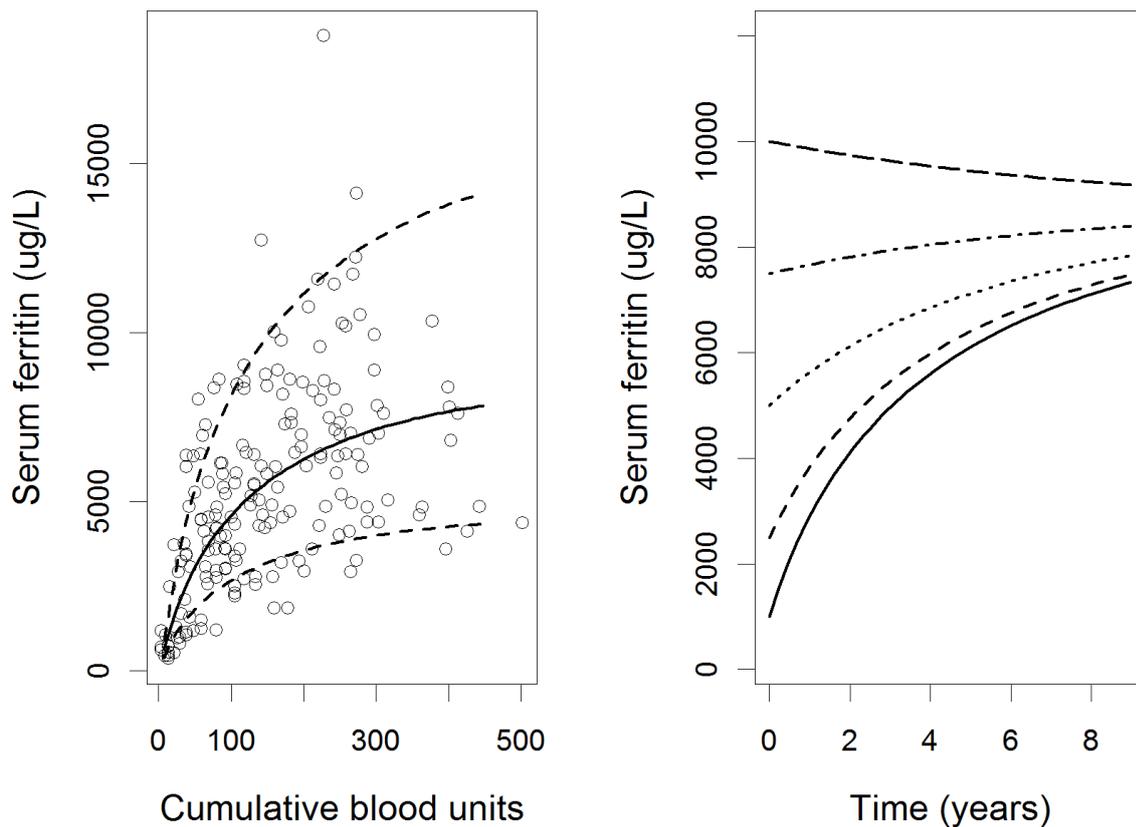


Figure 6. Visual predictive check of the disease model for iron overload and simulated mean serum ferritin profile in virtual patients not receiving iron chelation therapy. Left panel: VPC of the disease model for iron overload. Observed data are plotted using open circles; the black solid line represents the median of the simulated data; the dashed lines represent the 5th and 95th percentiles of the simulated data. Right panel: simulated ferritin profiles over a period of 10 years for a virtual patient not receiving iron chelation therapy. Each line represents a different ferritin baseline level: the solid, dashed (small), dotted, dashed-dotted, and dashed (big) lines represent 1000, 2500, 5000, 7500, and 10000 ug/L baseline ferritin levels.

6.4 Discussion and Conclusion

With this work we attempt for the first time to use a model-based approach in the field of transfusion-dependent diseases. An indirect response model was first developed in healthy subjects to account for the basal turnover of serum ferritin. Subsequently, the relationship between serum ferritin and cumulative blood units was quantified and integrated in the turnover model, and the non-linearity of the system was properly captured. Once the effect of the chronic transfusion regimen is introduced in the model, the contribution of the basal turnover of serum ferritin becomes negligible; the conversion rate (CRT) becomes the driving force of the changes in serum ferritin levels and gives a clear idea of the magnitude of iron overload in the absence of chelation therapy (Figure 6, right panel). As depicted in the same figure, the model allows exploring the natural course of the disease without treatment intervention; without such a model it would not be possible to appropriately quantify the true effect of iron chelation therapy. In addition, the nature of the model allows evaluating the drug effect of any available or future chelating agent.

Limitations

The lack of access to individual data did not allow a proper characterisation of the inter-individual variability and/or of a thorough covariate analysis. On the contrary, we could appropriately quantify the mean population changes in disease progression.

When integrating the Emax model with the turnover model, we assumed a constant time interval between subsequent units of blood transfused; the interval chosen was based on available literature data^{2,4,5,7,8}. Even though there is inter- and intra-patient variation in the transfusion regimen, we believe that the literature data support our assumption given that we could only evaluate the mean population profile of the integrated model.

Conclusions

In conclusion, despite some limitations due to incomplete availability of data a disease model was successfully developed in patients affected by severe iron overload that were not undergoing iron chelation therapy. The impact of blood transfusions on serum ferritin levels was quantified allowing a more mechanistic interpretation of the underlying disease progression. This model provides the basis for a more structured evaluation of therapeutic intervention in this patient population and gives the opportunity for further evaluation of the disease and its progression.

References

1. Barman Balfour J a, Foster RH. Deferiprone: a review of its clinical potential in iron overload in beta-thalassaemia major and other transfusion-dependent diseases. *Drugs*. 1999;58(3):553–78.
2. Galanello R, Origa R. Beta-thalassemia. *Orphanet J. Rare Dis*. 2010;5:11.
3. Ginzburg Y, Rivella S. B-Thalassemia: a Model for Elucidating the Dynamic Regulation of Ineffective Erythropoiesis and Iron Metabolism. *Blood*. 2011;118(16):4321–30.
4. Rebullà P. Blood transfusion in beta thalassaemia major. *Transfus. Med*. 1995;5(4):247–58.
5. Rebullà P, Modell B. Transfusion requirements and effects in patients with thalassaemia major. *Lancet*. 1991;337:277–280.
6. Rund D, Rachmilewitz E. Beta-Thalassemia. *N. Engl. J. Med*. 2005;353:1135–1146.
7. Thalassaemia International Federation (TIF). Guidelines for the clinical management of thalassaemia. 2008;
8. Porter J, Huehns E. Transfusion and exchange transfusion in sickle cell anaemias, with particular reference to iron metabolism. *Acta Haematol*. 1987;78:198–205.
9. Iancu T. Ferritin and hemosiderin in pathological tissues. *Electron Microscop Rev*. 1992;5:209–229.
10. Andrews NC, Schmidt PJ. Iron homeostasis. *Annu. Rev. Physiol*. 2007;69:69–85.
11. Anderson GJ, Frazer DM, McLaren GD. Iron absorption and metabolism. *Curr. Opin. Gastroenterol*. 2009;25(2):129–35.
12. Cazzola M, Huebers H, Sayers M, et al. Transferrin saturation, plasma iron turnover, and ferritin uptake in normal humans. *Blood*. 2011;66:935–939.
13. Cook JD, Barry WE, Hershko C, Fillet G, Finch C. Iron Kinetics with emphasis on iron overload. *Am. J. Pathol*. 1972;337–344.
14. Knovich MA, Storey JA, Coffman LG, Torti S V. Ferritin for the Clinician. *Blood Rev*. 2010;23(3):95–104.
15. Halliday JW, Powell LW. Ferritin and cellular iron metabolism. *Ann. N. Y. Acad. Sci*. 1988; 526:101–12.
16. Andrews N. Disorders of Iron Metabolism. *N. Engl. J. Med*. 1999; 341(26):1986–1995.

17. Borgna-Pignatti C, Cappellini MD, De Stefano P, et al. Survival and complications in thalassemia. *Ann. N. Y. Acad. Sci.* 2005; 1054:40–7.
18. Borgna-Pignatti C, Rugolotto S, De Stefano P, et al. Survival and complications in patients with thalassemia major treated with transfusion and deferoxamine. *Haematologica.* 2004; 89(10):1187–1193.
19. Borgna-Pignatti C, Cappellini MD, Stefano P De, et al. Cardiac morbidity and mortality in deferoxamine- or deferiprone-treated patients with thalassemia major. 2009;107(9):3733–3737.
20. Cunningham MJ, Macklin E a, Neufeld EJ, Cohen AR. Complications of beta-thalassemia major in North America. *Blood.* 2004;104(1):34–9.
21. Hershko C. Pathogenesis and management of iron toxicity in thalassemia. *Ann. N. Y. Acad. Sci.* 2010;1202:1–9.
22. Cappellini MD, Pattoneri P. Oral iron chelators. *Annu. Rev. Med.* 2009; 60:25–38.
23. Kwiatkowski JL. Oral iron chelators. *Pediatr. Clin. North Am.* 2008; 55(2):461–82.
24. Musallam KM, Taher AT. Iron chelation therapy for transfusional iron overload: a swift evolution. *Hemoglobin.* 2011; 35(5-6):565–73.
25. Shander A, Sweeney J. Overview of Current Treatment Regimens in Iron Chelation Therapy. *US Hematol.* 2009; 2(1):56–59.
26. Theil EC. Mining ferritin iron: 2 pathways. *Blood.* 2009;114(20):4325–6.
27. Brittenham GM, Cohen a R, McLaren CE, et al. Hepatic iron stores and plasma ferritin concentration in patients with sickle cell anemia and thalassemia major. *Am. J. Hematol.* 1993; 42(1):81–5.
28. Lipschitz D, Cook J, Finch C. A clinical evaluation of serum ferritin as an inndex of iron stores. *N. Engl. J. Med.* 1974; 290(22):1213–16.
29. Angelucci E, Brittenham G, McLaren C, et al. Hepatic iron concentration and total body iron stores in thalassemia major. *N. Engl. J. Med.* 2000; 343(5):327–331.
30. Villeneuve JP, Bilodeau M, Lepage R, Côté J, Lefebvre M. Variability in hepatic iron concentration measurement from needle-biopsy specimens. *J. Hepatol.* 1996; 25(2):172–7.
31. Fischer R, Longo F, Nielsen P, et al. Monitoring long-term efficacy of iron chelation therapy by deferiprone and desferrioxamine in patients with beta-thalassaemia major: application of SQUID biomagnetic liver susceptometry. *Br. J. Haematol.* 2003; 121(6):938–48.

32. Kirk P, Roughton M, Porter JB, et al. Cardiac T2* magnetic resonance for prediction of cardiac complications in thalassemia major. *Circulation*. 2009; 120(20):1961–8.
33. St Pierre TG, Clark PR, Chua-anusorn W, et al. Noninvasive measurement and imaging of liver iron concentrations using proton magnetic resonance. *Blood*. 2005; 105(2):855–61.
34. Dawkins S, Cavill I, Ricketts C, Worwood M. Variability of serum ferritin concentration in normal subjects. *Clin. Lab. Haematol*. 1979; 1(1):41–6.
35. Letsky E a, Miller F, Worwood M, Flynn DM. Serum ferritin in children with thalassaemia regularly transfused. *J. Clin. Pathol*. 1974; 27(8):652–5.
36. Poisot T. The digitize package: extracting numerical data from scatterplots. *The R Journal* Vol. 3/1. 2011; 25–26.
37. Comets E, Brendel K, Mentré F. Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. *Comput. Methods Programs Biomed*. 2008; 90(2):154–66.
38. Hooker AC, Staats CE, Karlsson MO. Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method. *Pharm. Res*. 2007; 24(12):2187–97.
39. Lindbom L, Ribbing J, Jonsson EN. Perl-speaks-NONMEM (PsN)--a Perl module for NONMEM related programming. *Comput. Methods Programs Biomed*. 2004; 75(2):85–94.
40. Worwood M, Cragg SJ, Jacobs a, et al. Binding of serum ferritin to concanavalin A: patients with homozygous beta thalassaemia and transfusional iron overload. *Br. J. Haematol*. 1980; 46(3):409–16.
41. George E, Wong H, George R, Ariffin W. Serum ferritin concentrations in transfusion dependent beta-thalassaemia. *Singapore med J*. 1994; 35:62–64.

Appendix

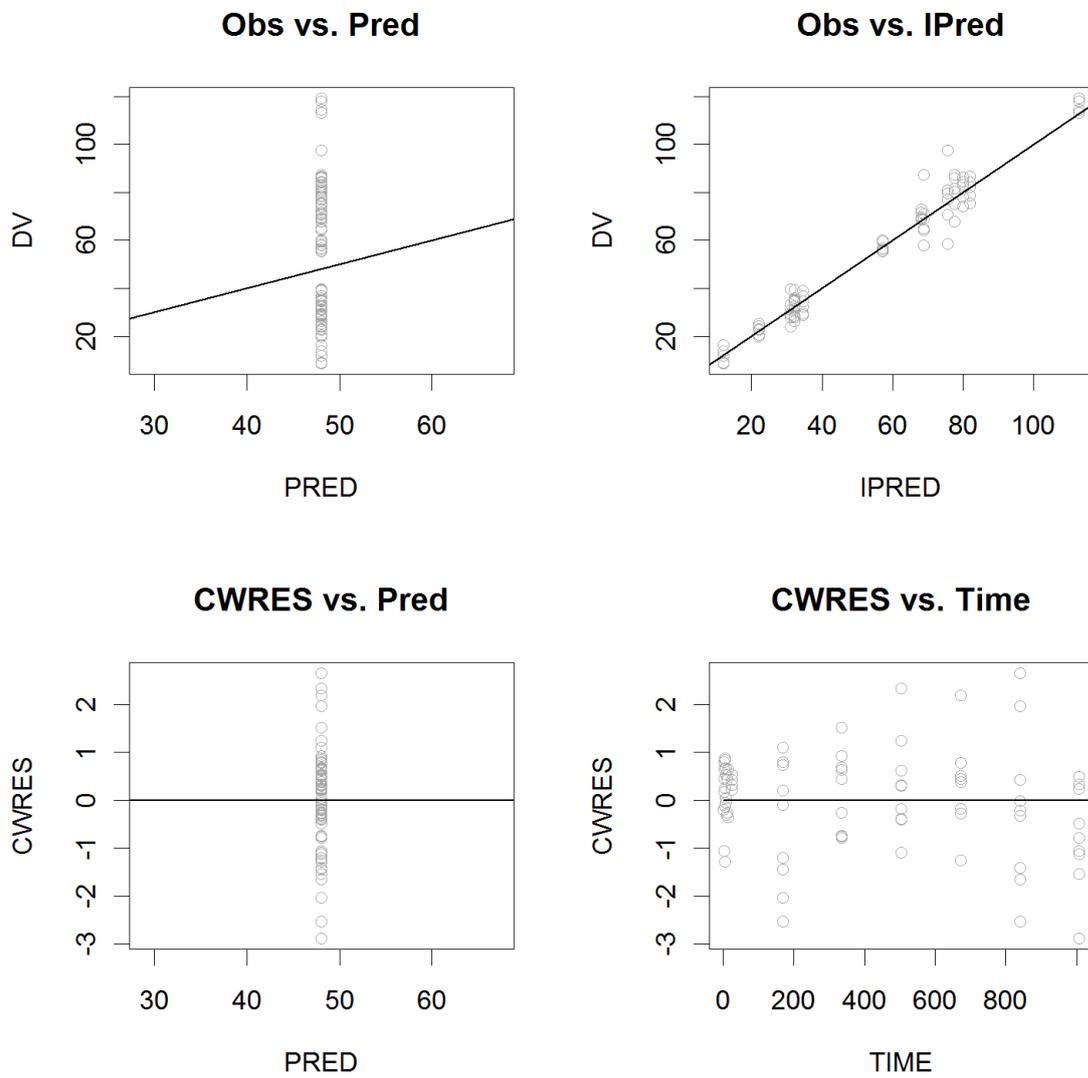


Figure 7. Goodness-of-fit plots for the turnover model in healthy individuals. Upper panels show the observed data (Obs) vs. population predictions (Pred) (left) and the observed data vs. individual predictions (IPred) (right). Lower panels show the conditional weighted residuals (CWRES) vs. population predictions (left) and the CWRES vs time (right).

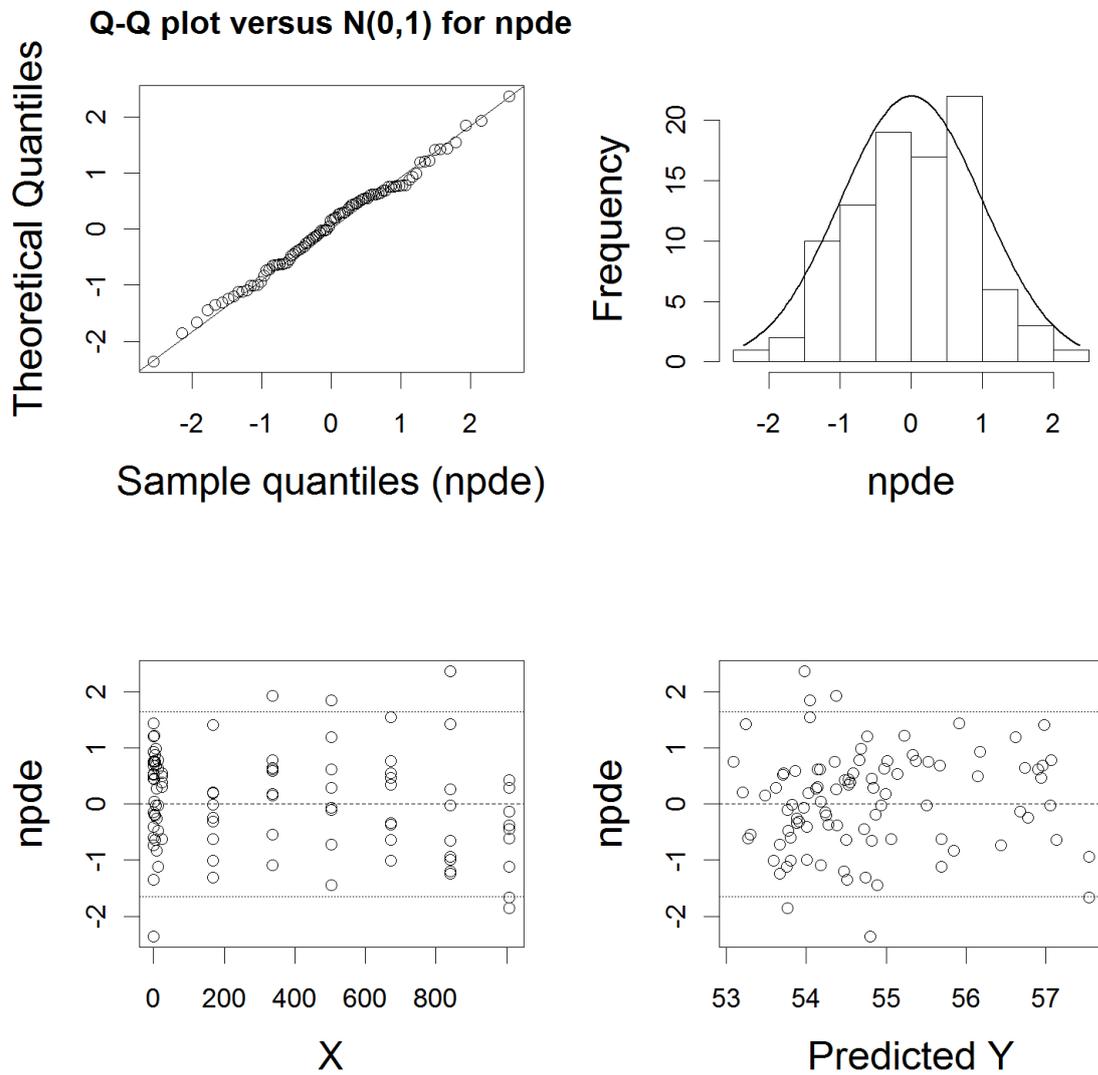


Figure 8. NPDE summaries for the turnover model in healthy individuals. Upper panels show the QQ-plot of the distribution of the NPDEs for a theoretical $N(0, 1)$ distribution (left) and the histogram of the distribution of the NPDE together with the density of the standard normal distribution (right). Lower panels show the NPDEs vs. time (left) and NPDEs vs. individual predictions (right).

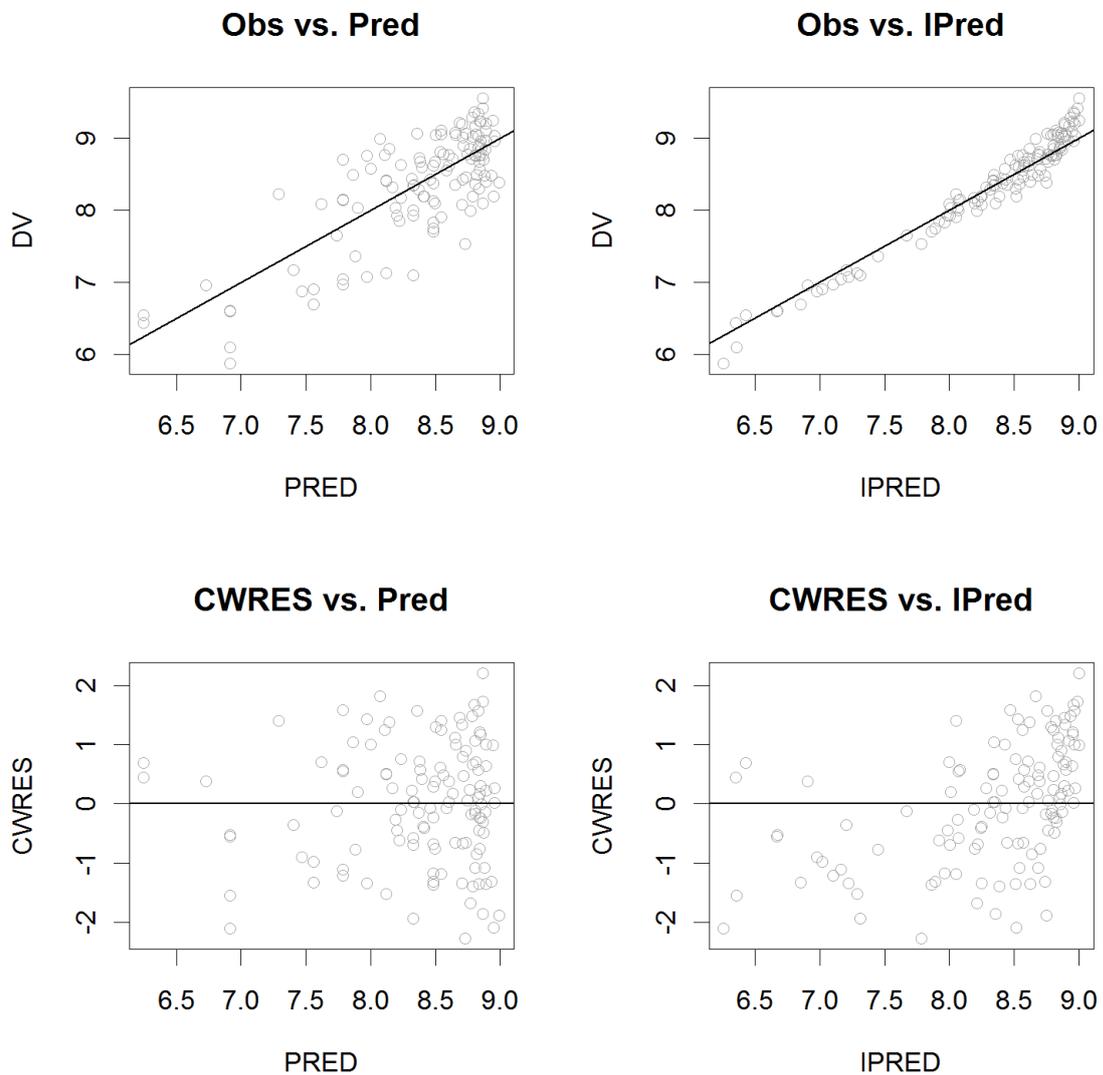


Figure 9. Goodness-of-Fit plots of estimated relationship between cumulative blood units and serum ferritin levels in patients affected by transfusion-dependent diseases not receiving iron chelation therapy: model building. Upper panels show the observed data (Obs) vs. population predictions (Pred) (left) and the observed data vs. individual predictions (IPred) (right). Lower panels show the conditional weighted residuals (CWRES) vs. population predictions (left) and the CWRES vs IPred (right).

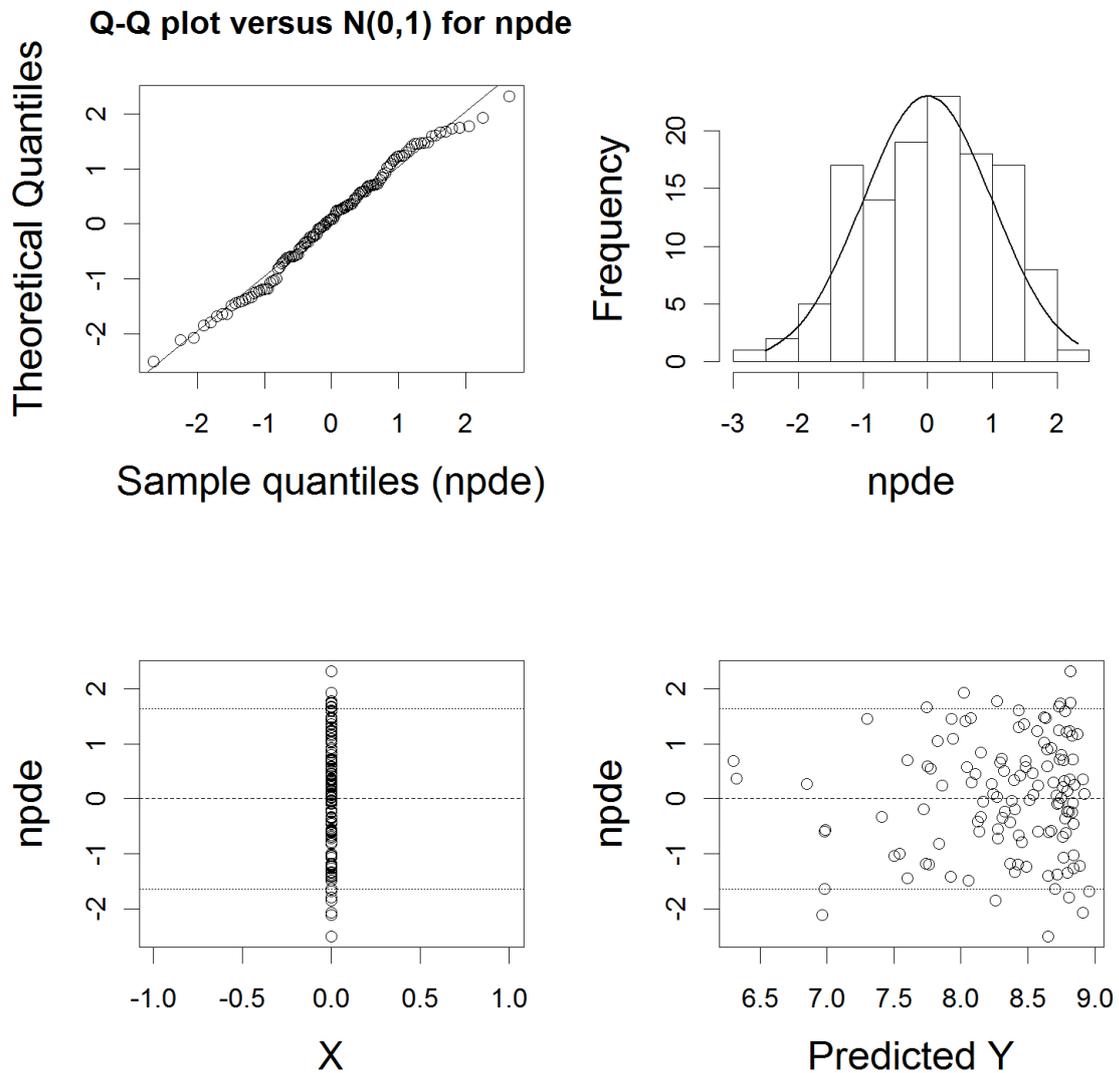


Figure 10. NPDE summaries of estimated relationship between cumulative blood units and serum ferritin levels in patients affected by transfusion-dependent diseases not receiving iron chelation therapy: model building. Upper panels show the QQ-plot of the distribution of the NPDEs for a theoretical $N(0, 1)$ distribution (left) and the histogram of the distribution of the NPDE together with the density of the standard normal distribution (right). Lower panels show the NPDEs vs. time (left) and NPDEs vs. individual predictions (right).

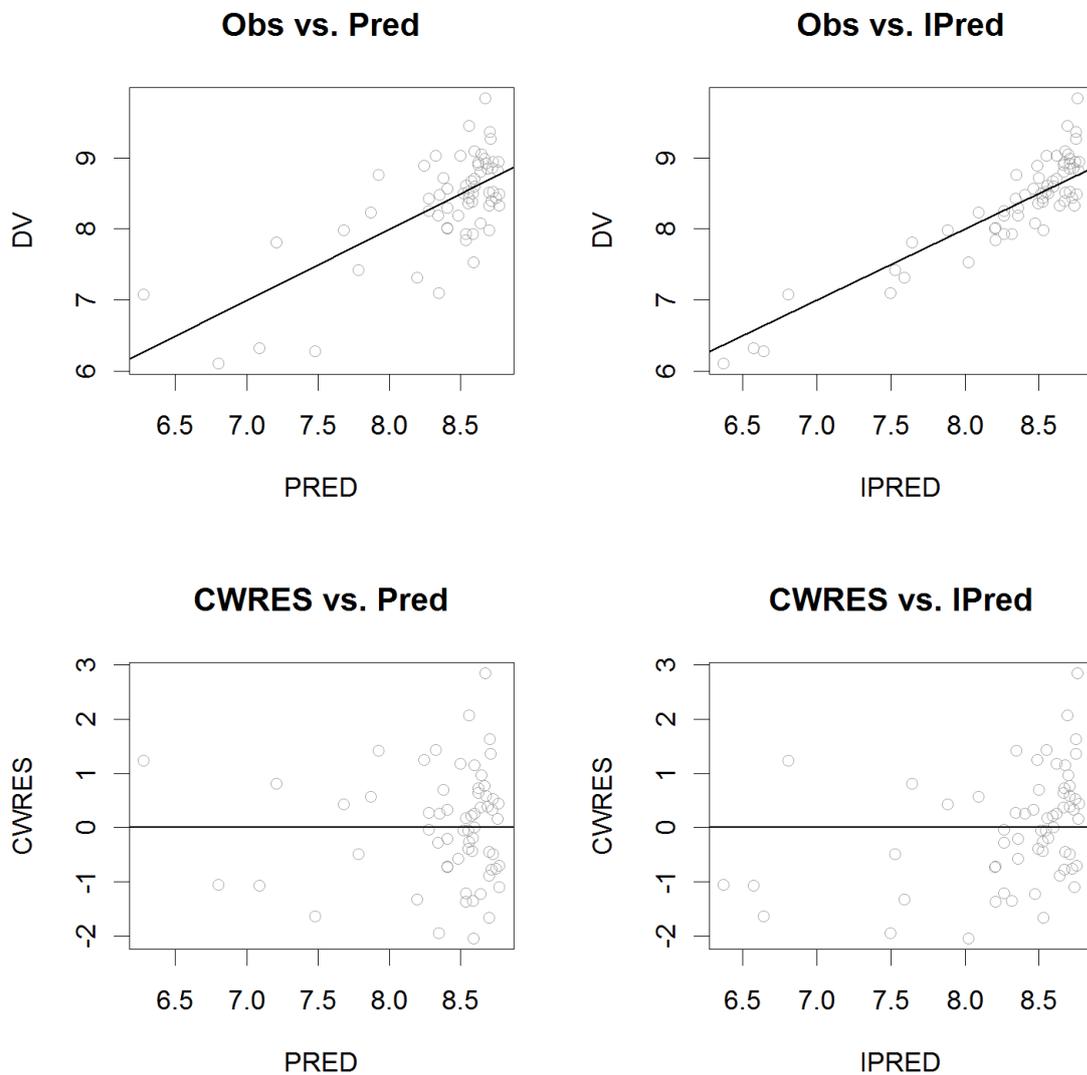


Figure 11. Goodness-of-Fit plots of estimated relationship between cumulative blood units and serum ferritin levels in patients affected by transfusion-dependent diseases not receiving iron chelation therapy: external validation. Upper panels show the observed data (Obs) vs. population predictions (Pred) (left) and the observed data vs. individual predictions (IPred) (right). Lower panels show the conditional weighted residuals (CWRES) vs. population predictions (left) and the CWRES vs IPred (right).

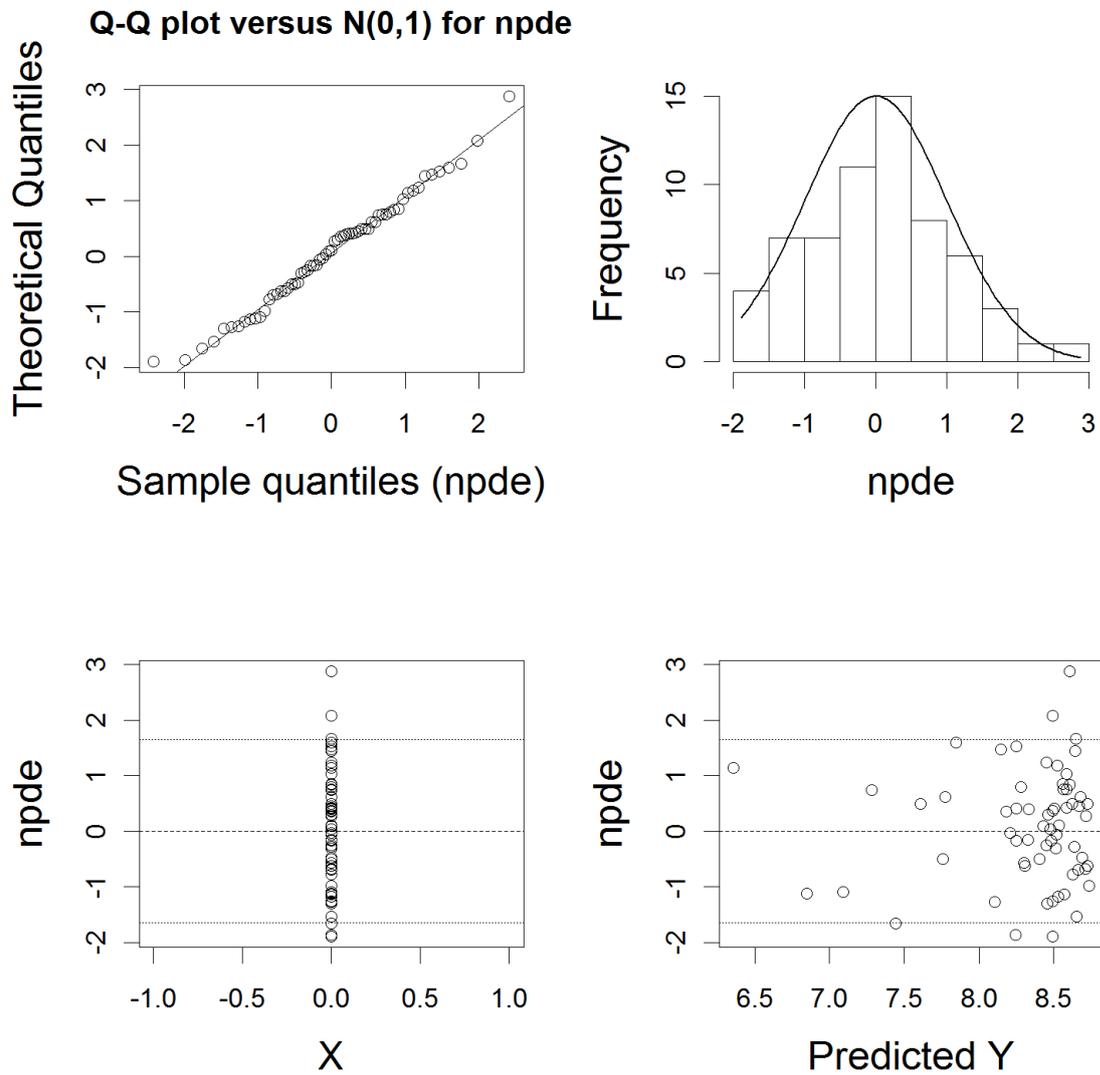


Figure 12. NPDE summaries of estimated relationship between cumulative blood units and serum ferritin levels in patients affected by transfusion-dependent diseases not receiving iron chelation therapy: external validation. Upper panels show the QQ-plot of the distribution of the NPDEs for a theoretical $N(0, 1)$ distribution (left) and the histogram of the distribution of the NPDE together with the density of the standard normal distribution (right). Lower panels show the NPDEs vs. time (left) and NPDEs vs. individual predictions (right).

CHAPTER 7

Model-based optimisation of deferoxamine chelation therapy

Francesco Bellanti, Giovanni C. Del Vecchio, Maria C. Putti, Carlo Cosmi, Ilaria Fotzi, Suruchi D. Bakshi, Meindert Danhof, and Oscar Della Pasqua

Pharm Res – In press

Summary

Purpose: Our endeavour is to show the advantages of a model-based approach to identify key factors that play a role in transfusion-dependent iron overload. We use deferoxamine as a paradigm compound to assess the role of relevant covariates on the underlying disease progression.

Methods: Data from clinical routine practice on 27 patients affected by β -thalassaemia major were used for the analysis. Serum ferritin was selected as the main endpoint of interest for the assessment of iron overload. Its time course was characterised by means of a hierarchical nonlinear mixed effects model, as implemented in NONMEM (7.2.0).

Results: A turnover model best described serum ferritin changes over time, with the effect of blood transfusions introduced as a change in the ferritin conversion rate, whereas the effect of deferoxamine was described by a proportional change in the degradation rate constant (K_{out}). The inclusion of IOV (57.4 %) on the conversion rate resulted in a significant drop in the OFV (Δ 443) allowing a better description of the individual profiles.

Conclusions: A model-based approach was successfully used to gather further insight into the dynamics of ferritin in transfusion-dependent iron overload. Given the choice of parameterisation, the model may be used as a tool to support clinical practice, including the evaluation of the dose rationale for existing and novel chelating agents.

Abbreviations

ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
C _{ss}	Steady state concentration
CMPL	Treatment compliance
DFO	Deferoxamine
FT ₄	Free T ₄
IIV	Inter-individual variability
IOV	Inter-occasion variability
M&S	Modelling and simulations
NPDE	Normalised predictive distribution error
NTBI	Non-Transferrin Bound Iron
OFV	Objective function value
PKPD	Pharmacokinetic-pharmacodynamic
PRED	Population prediction
RBC	Red blood cells
TSH	Thyroid-Stimulating Hormone
VPC	visual predictive check

7.1 Introduction

Transfusional iron overload

Beta-thalassaemia major is a hereditary blood disorder and patients affected by this disease require regular red blood cell (RBC) transfusions to survive (1–7). Without the chronic transfusion regimen, patients would die before the third decade of life (2,4,5,7,8). Even though a significant improvement has been achieved in the management of the chronic transfusion regimen in the past decades, therapeutic intervention will eventually lead to a series of complications. Iron overload is the most common and relevant one and it is associated with several co-morbidities such as cardiac dysfunction, liver fibrosis, hypogonadism, hypothyroidism, hypoparathyroidism and diabetes mellitus (6,9,10). Cardiac disease caused by myocardial siderosis is the most relevant complication, causing death in 71% of the patients affected by transfusion-dependent diseases (11).

For complex processes such as iron overload, understanding of the dynamics of the disease and its progression is crucial to adequately evaluate the therapeutic intervention. This complexity is also characterised by the fact that the currently accepted biomarker, i.e., blood ferritin is not specific enough to distinguish the effect of transfusion from the influence of other pathological mechanisms such as inflammatory disorders, and/or liver status, which can equally affect the iron interchange between organs and the circulatory system (2,12–14). Consequently, ferritin levels may not provide a direct link for total body or tissue specific iron accumulation at specific time points. On the other hand, changes in ferritin levels over time are still helpful for the management of the disease and maintaining serum ferritin below a threshold of about 2500 µg/L is a widely accepted therapeutic goal (2,3,5–7). However, important clinical questions are not yet fully understood, e.g. how much time is required in order to observe a stable response or to reach clinically meaningful serum ferritin levels.

Iron chelation therapy with chelating agent deferoxamine

Given that human physiology does not have an innate mechanism that allows removal of the iron excess, treatment with iron chelators is therefore vital to prevent its accumulation and related complications (15–18). In the current investigation we attempt to characterise the (patho)-physiological response to the iron chelating agent deferoxamine as a paradigm compound for the assessment of iron dynamics using a model-based approach. Deferoxamine was the first iron chelator approved for human use and has been available for the treatment of iron overload for more than 35 years (2,6,15–19). It is an exadentate chelator that binds iron in a 1:1 ratio. The drug is rapidly absorbed after intramuscular and subcutaneous administration, but it cannot be absorbed orally. In the treatment of iron overload in patients affected by transfusion-dependent haemoglobinopathies several dosing regimens and dose levels have been proposed and used in the past but in the majority of

cases deferoxamine is given as an 8 to 12 hour nightly subcutaneous infusion (5 to 7 days a week) (2,19–21). The serum protein binding is less than 10% and the drug undergoes the following metabolic reactions: transamination and oxidation; beta-oxidation; decarboxylation and N-hydroxylation. The average recommended daily dose varies between 20 and 60 mg/kg and the drug has an half-life of 5.6 hours in patients (20–22).

Deferoxamine binds free iron by preventing the uptake of NTBI (Non-Transferrin Bound Iron) into organs but is also acts within the cell where enters by endocytosis, stimulates the degradation of ferritin via lysosomes and subsequently binds the released iron. The iron bound to deferoxamine is then excreted in urine and faeces (2,6,21,23).

Regardless of numerous limitations associated with the use of deferoxamine, such as poor compliance due to the parenteral administration, inadequate cardiac iron removal and auditory, ocular and neurological toxicities (6,16,18,19,24), deferoxamine is still the most common used therapy for the treatment of iron overload. This widespread use has remained despite the presence of new oral iron chelators.

Given the complexity of the issues highlighted above, our focus is to gain insight into key factors that play a role in iron overload; with the objective of quantifying the therapeutic effect of deferoxamine and identify potential covariates on model parameters describing the underlying disease progression. Furthermore, we propose how modelling and simulation (M&S) can be applied to support decision making in clinical practice, providing a framework to predict changes in the disease status and resulting ferritin response following treatment with existing and novel chelators.

7.2 Methods

Data

The modelling analysis was performed using retrospective clinical data from three different Italian centres: Azienda Ospedaliera Universitaria Consorziale Policlinico di Bari U.O. Pediatria Federico Vecchio; Azienda Ospedaliera Universitaria Policlinico di Sassari Clinica Pediatrica, ASL 1 D.H. per Talassemia; Azienda Ospedaliera di Padova Clinica di Oncoematologia Pediatrica. The study has been conducted in full conformance with the principles of the Declaration of Helsinki and with the local laws and regulations concerning clinical trials. The protocol and the informed consent documents have been formally approved by the relevant research ethics committee of each clinical site.

27 patients affected by transfusion-dependent diseases, receiving deferoxamine as single drug for iron chelation therapy were considered eligible for the retrospective study. Patients receiving chemotherapy, and/or affected by other diseases that require additional blood

transfusions, and/or affected by unrelated endocrine dysfunctions were considered not eligible for the study. Baseline characteristics of the patient population are provided in Table 1. Serum ferritin was the main endpoint of interest and was measured every two to three months; patients contributed with 40.2 observations on average (sd: 17), with a minimum of 4 samples per year. With the same frequency, data on the following variables were collected during the study: height and body weight of the patients, as well as clinical data on TSH, FT4, AST, ALT, glucose, creatinine and ejection fraction. These were considered as potential covariates and were tested during the study to evaluate their influence on the changes in serum ferritin levels.

Table 1. Baseline characteristics of the patient population

	Units	Median	Range
Age	Years	14.6	6.8-19.9
Weight	Kg	46	17.5-71
Height	Cm	154	111-173
TSH	mIU/L	2.34	0.58-83.2
FT4	ng/dL	1.05	0.73-1.43
AST	U/L	33	7-159
ALT	U/L	56	9-372
Glucose	mg/dL	91	52-444
Creatinine	mg/dL	0.6	0.2-1.12
Ejection Fraction	%	64	35-77
Ferritin	µg/L	2260	393-8500

PK model deferoxamine

As pharmacokinetic samples are not collected in clinical routine monitoring, the model was built using literature data (25) by fitting a mean pharmacokinetic profile in adults patients affected by transfusion-dependent haemoglobinopathies receiving deferoxamine as an 8 hours subcutaneous infusion. A two compartment pharmacokinetic model with zero-order absorption (8 hours subcutaneous infusion) and first-order elimination processes provided an appropriate description of the average steady state concentration (C_{ss}^{AV}) for the population of interest. The fitting of the published data is shown in Figure 1.

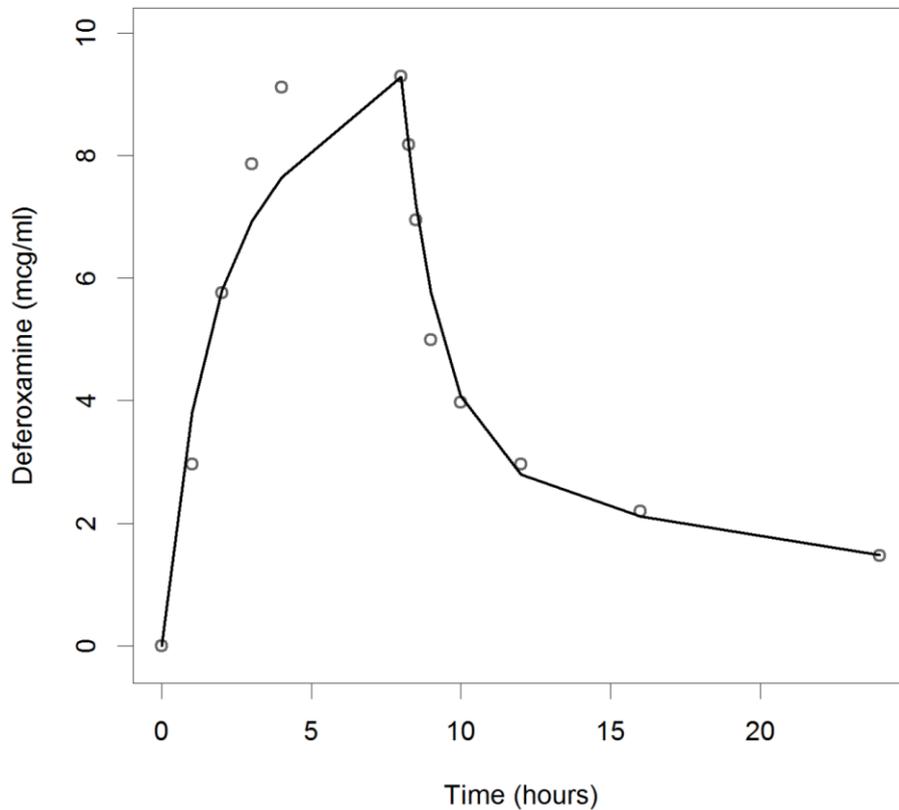


Figure 1. Performance of the pharmacokinetic model of deferoxamine. The circles represent the mean population deferoxamine concentrations reported in literature (23). The solid line represents model prediction.

Assumptions were then made to allow the use of the model to predict exposure in the patient population: 1) the simulations were based on the dosing regimen information collected in the clinical centres and the reported changes to the regimen; and 2) in the absence of quantitative data on adherence to drug therapy, compliance was assumed equivalent to 100%. The role of compliance was assessed in a second phase and details on that are provided in the next paragraphs. Afterwards, fixed allometric scaling (exponent of 0.75 on CL/F and 1.00 on V/F) was used to extrapolate C_{ss}^{AV} in adolescents and children. Population prediction (PRED) were used to generate C_{ss}^{AV} values in the population of interest.

Disease model iron overload

A disease model for iron overload in patients affected by transfusion-dependent diseases was previously developed by our group [unpublished results]. It consists of an indirect

response model where basal turnover of ferritin levels is depicted by a zero-order production rate (K_{in}) and a first-order degradation rate (K_{out}) and the disease component is described by an additive production rate (CRT) triggered by the transfusion regimen which was found to be non-linearly correlated to the disease status (actual ferritin levels).

$$\frac{dFERRITIN}{dt} = K_{in} + CRT - K_{out} \times FERRITIN \quad \text{Equation 1}$$

$$CRT = SCL \times e^{-SHP \times FERRITIN} \quad \text{Equation 2}$$

where SCL is a scaling factor and SHP is the shape factor of the correlation. The population parameters of the disease model were kept fixed when performing the PKPD analysis of the retrospective clinical data.

Modelling

The software R (v.2.14.0) was used for statistical summaries as well as data manipulation and preparation for modelling purposes. Nonlinear mixed effects modelling was instead performed in NONMEM version 7.2 (Icon Development Solutions, USA).

Model building criteria included: (i) successful minimisation, (ii) standard error of estimates, (iii) number of significant digits, (iv) termination of the covariance step, (v) correlation between model parameters and (vi) acceptable gradients at the last iteration. Comparison of hierarchical models was based on the likelihood ratio test. Goodness of fit was assessed by graphical methods, including population and individual predicted vs. observed concentrations, conditional weighted residual vs. observed concentrations and time, correlation matrix for fixed vs. random effects, correlation matrix between parameters and covariates and normalised predictive distribution error (NPDE) (26).

Fixed and random effects were introduced into the model in a stepwise manner. Inter-individual variability in the parameters was assumed to be log-normally distributed. A parameter value of an individual i (post hoc value) is therefore given by the following equation:

$$\theta_i = \theta_{TV} * e^{\eta_i}$$

in which θ_{TV} is the typical value of the parameter in the population and η_i is assumed to be random variable with zero mean and variance ω^2 . Residual variability, which comprises measurement and model error, was described with a proportional error model. This means for the j th observed concentration of the i th individual, the relation Y_{ij} :

$$Y_{ij} = F_{ij} + \epsilon_{ij} * W$$

where F_{ij} is the predicted concentration and ε_{ij} the random variable with mean zero and variance σ^2 . W is a proportional weighing factor for ε .

Different concentration-effect relationships (e.g., Emax model, linear model, etc.) were tested on the disease model presented in equation 1 to quantify the effect of deferoxamine on serum ferritin levels. Css^{AV} levels, generated with the PK model described above were used in the drug model as a measure of deferoxamine exposure. The effect of deferoxamine (DFO) was introduced as proportional change in the degradation rate ($Kout$) of ferritin as shown in equation 3 which is derived from equation 1.

$$\frac{dFERRITIN}{dt} = Kin + CRT - Kout \times FERRITIN \times (1 + DFO) \quad \text{Equation 3}$$

$$DFO = SLP \times SCss^{AV} \quad \text{Equation 4}$$

where DFO is the effect of deferoxamine on the $Kout$ of the disease model, SLP is the slope parameter of the concentration-effect relationship, and $SCss^{AV}$ is the steady state concentrations simulated with the deferoxamine PK model.

In addition, the disease model parameters, the scaling (SCL) and the shape (SHP) factors presented in equation 2 were found to be non-linearly correlated to the actual disease status according to the following relationships:

$$SCL_i = SCL_{ref} \times \left(\frac{FERRITIN}{FERRITIN_{med}} \right)^{\theta x} \quad \text{Equation 5}$$

$$SHP_i = SHP_{ref} \times \left(\frac{FERRITIN}{FERRITIN_{med}} \right)^{\theta x} \quad \text{Equation 6}$$

where SCL_{ref} and SHP_{ref} are the reference parameters in the population of interest, SCL_i and SHP_i are the individual parameters, $FERRITIN_{med}$ is the median ferritin value in the population of interest and θx is the estimated exponent of the relationship.

The evaluation of the final model was based on graphical and statistical methods, including visual predictive checks (27). Bootstrap was used to identify bias, stability and accuracy of the parameter estimates (standard errors and confidence intervals). The bootstrap procedures were performed in PsN v3.5.3 (University of Uppsala, Sweden) (28), which automatically generates a series of new data sets by sampling individuals with replacement from the original data pool, fitting the model to each new data set.

Assessing the role of compliance

At the initial stage of the model-building phase, the model was not able to appropriately describe the data (Figure 2), under the assumption that the patients represent a single population. However, two different profiles were observed in the data, which prompted us consider dichotomising the data into responders and non-responders. An arbitrary definition was used based on the observed ferritin levels: the responders showed very stable profiles around 2500 $\mu\text{g/L}$ serum ferritin, whereas the non-responders showed very steep increases in ferritin levels and appeared not to be able to return to a less severe state of the disease. A mixture model improved the quality of the fitting, but did not allow an adequate characterisation of the individual profiles.

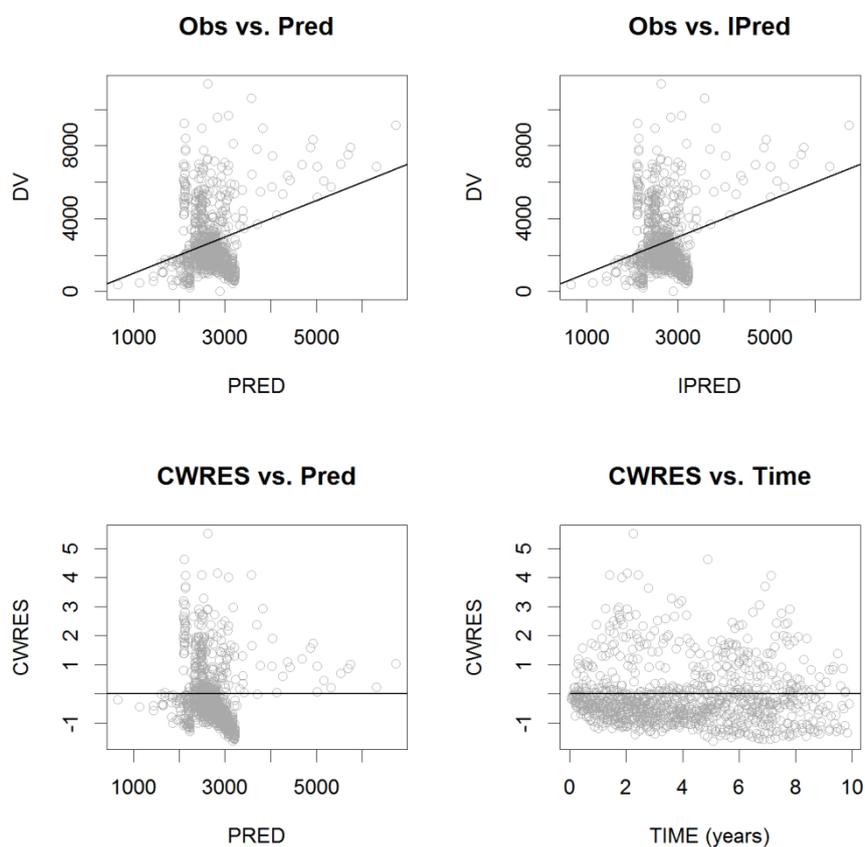


Figure 2. Goodness-of-fit plots of the model without the inclusion of compliance. Upper panels show the observed data (Obs) vs. population predictions (Pred) (left) and the observed data vs. individual predictions (IPred) (right). Lower panels show the conditional weighted residuals (CWRES) vs. population predictions (left) and the CWRES vs time (left).

We have therefore investigated other potential mechanisms and explanatory factors associated with the different response profiles based on the information available in the

literature. Treatment compliance was found to be the major cause of differences in ferritin levels. In the papers by Gabutti et al. (29) and Galanello et al. (30) serum ferritin profiles are quite stable over the observational period (Figures 3 and 1 respectively) as in our responder group and compliance is in both cases higher than 95%. In other investigations (20,29,31) Kaplan-Meier analyses show the relationship between survival and treatment compliance, providing evidence of the fact that poor adherence has a crucial impact on the clinical outcome. In particular the work by Olivieri et al. (31) shows how survival can directly be linked to the observed ferritin levels.

The absence of quantitative data on treatment compliance in our retrospective study did not allow us to directly select this variable to account for such differences, which represented a clear obstacle for the analysis. To overcome this issue we used the work carried out by Olivieri et al. (31) as a reference and we derived a new variable (CMPL) based on the percentage of observations for each patient above the threshold of 2500 µg/L ferritin. The new variable (CMPL) was introduced in the model as follows:

$$DFO = SLP \times TC_{ss}^{AV} \quad \text{Equation 7}$$

$$TC_{ss}^{AV} = SC_{ss}^{AV} \times (1 - CMPL) \quad \text{Equation 8}$$

where DFO is the effect of deferoxamine, SLP is the slope parameter of the concentration-effect relationship, and TC_{ss}^{AV} are the “true” steady state concentrations after accounting for the impact of treatment compliance (CMPL). TC_{ss}^{AV} are derived from the simulated steady state concentrations (SC_{ss}^{AV}) corrected by treatment compliance as shown in equation 7.

The implementation of treatment compliance provided a significant increase in the fitting performances of the model (as shown in Figures 3 and 4) and allowed a more accurate quantification of the therapeutic intervention.

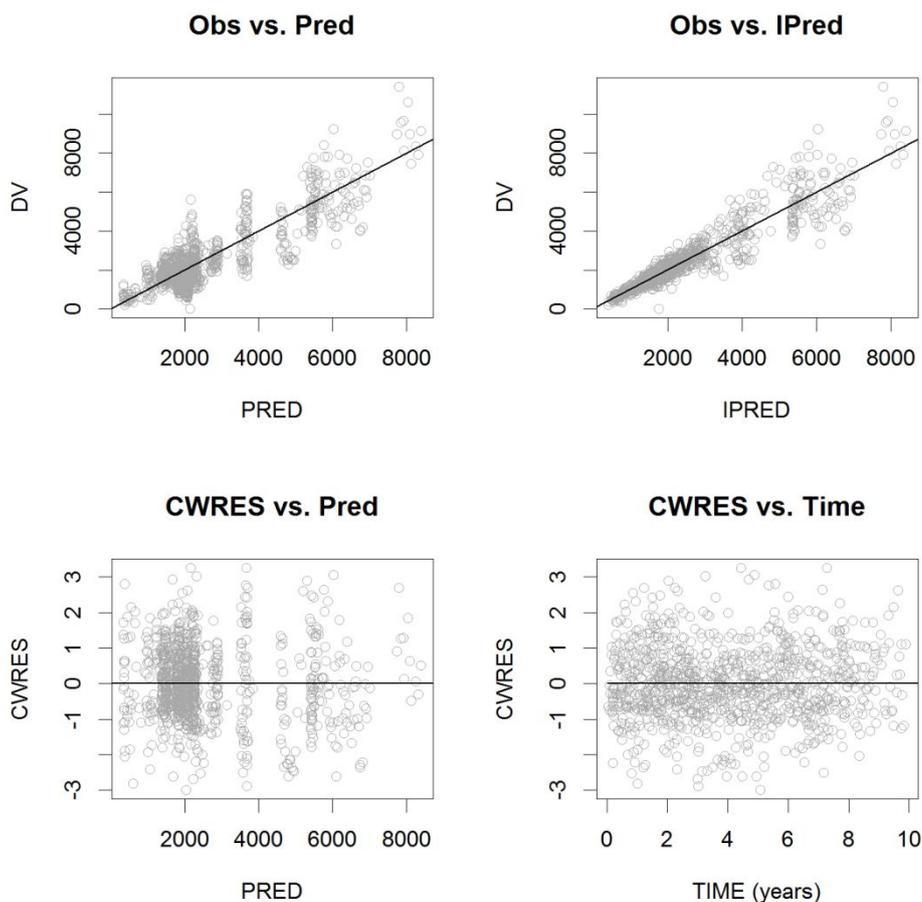


Figure 3. Goodness-of-fit plots of the final model. Upper panels show the observed data (Obs) vs. population predictions (Pred) (left) and the observed data vs. individual predictions (IPred) (right). Lower panels show the conditional weighted residuals (CWRES) vs. population predictions (left) and the CWRES vs time (left).

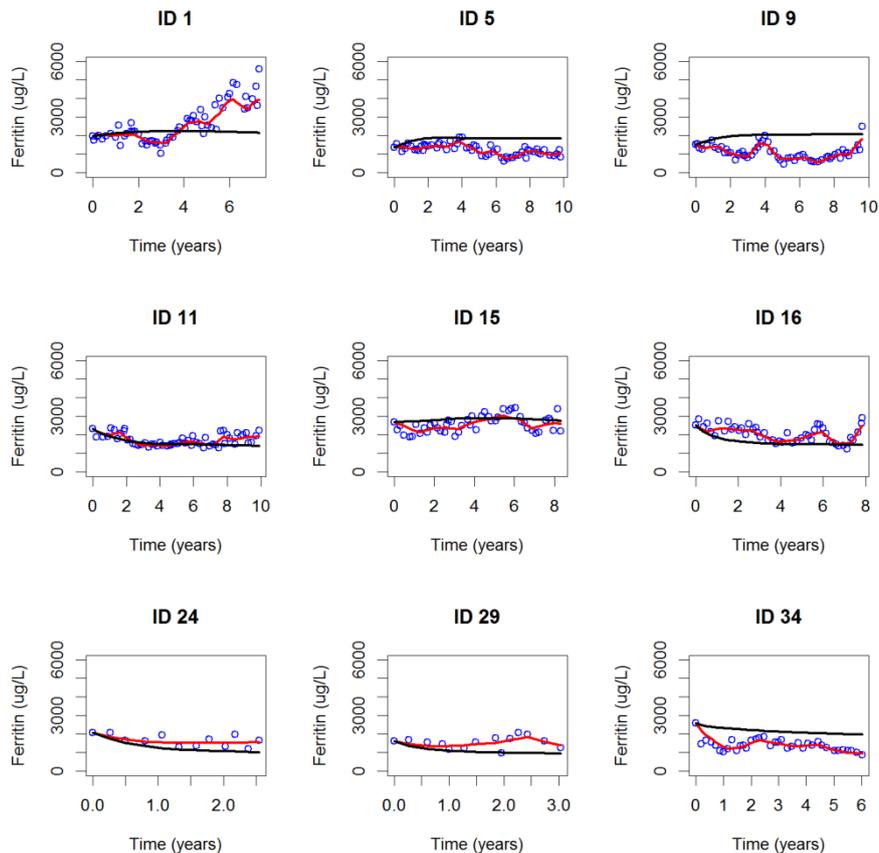


Figure 4. Individual plots of 9 randomly selected patients: observed data are plotted using blue circles; the black solid line represents the population prediction (Pred) and the red solid line represents the individual predictions (IPred).

Model Simulations

Simulations were performed to investigate the impact of different exposure levels and various compliance scenarios on the clinical evaluation of serum ferritin levels. Time to reach 2500 $\mu\text{g/L}$ serum ferritin (threshold between moderate and severe iron overload (2,6,20,31,32)) was chosen as a comparison measure between different scenarios. The differential equation solver ode15s from the software MATLAB (version R2010b) was used for the purpose of model simulations, whereas the software R (v.2.14.0) was used for graphical summaries. The ode15s which is a multistep solver and uses numerical differentiation formulas is particularly suitable for stiff systems (33,34).

Three dosing regimens (30, 45 and 60 mg/kg/day for 5 days a week) were used to generate C_{ss}^{AV} in a patient population with body weight ranging from 15 to 75 kg. This allowed evaluating the impact of different exposures on the endpoint of interest (time to reach the threshold). Each exposure level was then tested on patients starting at different baselines

ranging from 3000 to 12000 $\mu\text{g/L}$ serum ferritin. In this set of simulations the exposure was assumed to be constant over time.

In addition, simulations were used to evaluate the impact of various compliance scenarios on changes in ferritin levels. Our main interest in this case was to show how the model can be used prospectively in the clinical practice to evaluate a priori any given situation. To present and discuss the results of these simulations, given the large number of scenarios evaluated, we considered only a virtual patient of 45 kg receiving 45 mg/kg/day deferoxamine 5 days per week (representative of the mean patient in the population under investigation). The different scenarios investigated are presented in Table 2; compliance is stratified per 1 year, 6, 2 and 1 months.

Table 2. Simulation scenarios for the evaluation of different compliance levels

		Number of missed doses in the stratification period				
		Single doses (Random)	Consecutive doses (Drug holidays)			
			Stratification			
	% of missed doses	1 year	1 year	6 months	2 months	1 month
Scenario 1	10%	25	25	/	5	/
Scenario 2	20%	50	50	25	10	5
Scenario 3	30%	75	75	/	15	/
Scenario 4	40%	100	100	50	20	10
Scenario 5	50%	125	125	/	25	/
Scenario 6	60%	150	150	75	30	15
Scenario 7	70%	175	175	/	35	/
Scenario 8	80%	200	200	100	40	20
Scenario 9	90%	225	225	/	45	/

Full adherence is equivalent to 250 doses per year

The iterations were stopped if more than 5 years were needed to reach the threshold of 2500 $\mu\text{g/L}$ serum ferritin. As proposed in the work carried out by Piana et al. (35), the five scenarios selected try to cover different compliance patterns that may occur in the presence of a chronic regimen. Scenario 1 with single doses missed at random reflects poor quality of execution, whereas the other scenarios provide a range of options that reflect different patterns and durations of drug holidays.

7.3 Results

Disease model

The use of a disease model describing the impact of blood transfusions on serum ferritin was expanded to include the effects of chelation therapy. The effect of blood transfusions was introduced as a conversion rate on the production rate of ferritin and was found in the previous analysis to be inversely correlated to the disease status as shown in equation 1 and 2. In addition, the disease model parameters, the scaling (SCL) and the shape (SHP) factors presented in equation 2 were found to be non-linearly correlated to the actual disease status. Their inclusion in the model provided a significant decrease in the objective function value (OFV) and allowed a better description of the data. Furthermore, the inclusion of inter-occasion variability (IOV = 57.4 %) on the conversion rate resulted in a significant drop in the OFV (Δ 443) allowing a better description of the individual profiles.

Drug model

The effect of deferoxamine (DFO) was introduced in a proportional way on the degradation rate (K_{out}) of ferritin. Furthermore, the implementation of treatment compliance as a factor on the exposure of deferoxamine improved considerably the data fitting and the model performance, as well as the inclusion of inter-individual variability (IIV) on the slope parameter, which reduced significantly the OFV and improved goodness of fit and visual predictive check (VPC) diagnostics. An overview of the final model parameters and bootstrap results is presented in Table 3.

Table 3. Parameter estimates of the PKPD model of deferoxamine

Parameter	Estimate	Bootstrap (mean)	CV (%)
Kin (mcg/h)	0.0002 (FIX)	/	/
Kout (h-1)	0.0000045 (FIX)	/	/
SHP (h-1)	0.00026 (FIX)	/	/
SCL (mcg/h)	0.383 (FIX)	/	/
Slope (mcg/conc)	4.81	5.16	15.7
Error Proportional	-0.173	-0.17	6.5
DIS exp on SHP	1.29	1.08	57.4
DIS exp on SCL	0.845	0.67	51.9
IIV on Slope	0.082	0.105	80.9
IOV on CRT	0.252	0.29	43.1

Internal model validation diagnostics were satisfactory. Individual predicted profiles and goodness-of-fit plots as shown in Figures 3 and 4, as well as VPC (Figure 5) reveal that the

model provides an adequate and non-biased description of the data. In addition, NPDE summaries (Figure 6) show that the discrepancy between predicted and observed values can be assumed to be normally distributed.

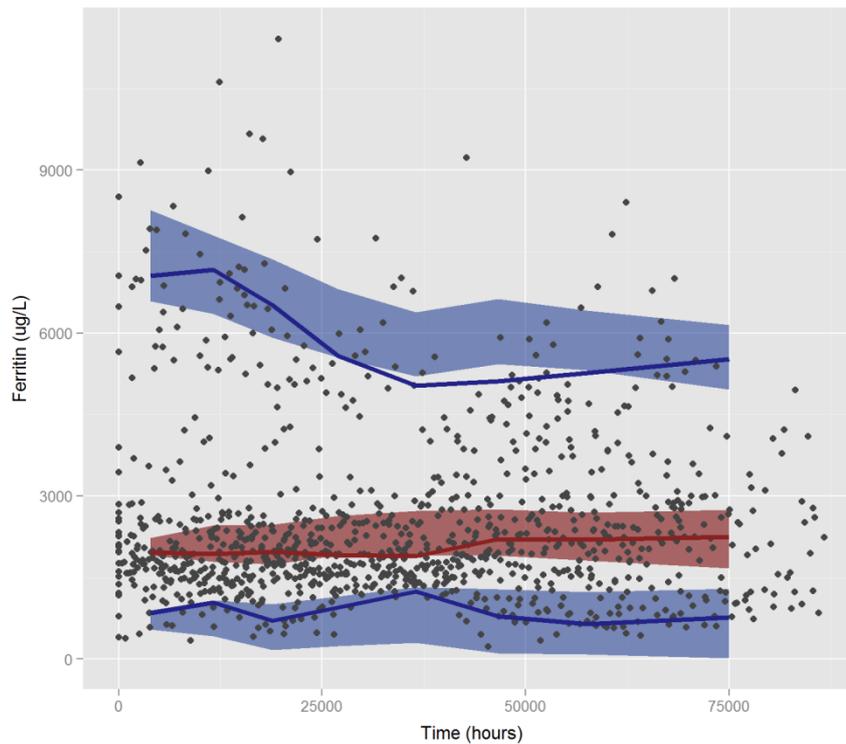


Figure 5. Visual predictive check: observed data are plotted using grey circles; the red solid line represents the median of the observed data; the blue solid lines represent the 5th and 95th percentiles of the observed data. The red shaded area represents the 95th CI of the median of the simulated data; the blue shaded areas represent the 95th CI of the 5th and 95th percentiles of the simulated data.

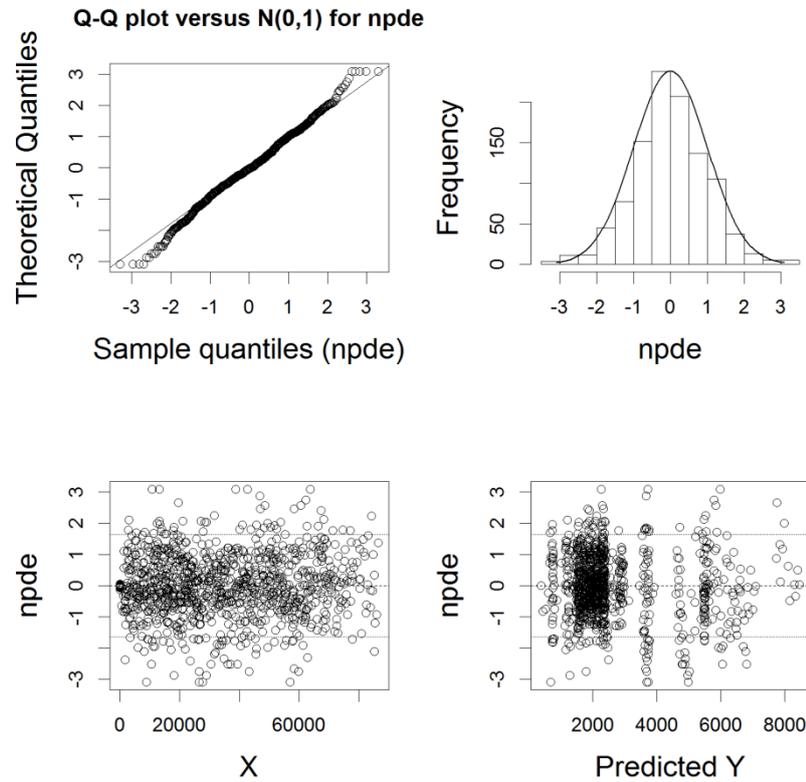


Figure 6. NPDE: normalised prediction distribution errors. Upper panels show the QQ-plot of the distribution of the NPDEs for a theoretical $N(0, 1)$ distribution (left) and the histogram of the distribution of the NPDE together with the density of the standard normal distribution (right). Lower panels show the NPDEs vs. time (left) and NPDEs vs. individual predictions (right).

Model Simulations

The impact of different exposure levels and various compliance scenarios on the clinical evaluation of serum ferritin levels was evaluated through model simulations. The results of the effect of different exposure levels in virtual patients characterised by different body weights and starting at different ferritin baseline levels are presented in Figure 7 for the following dosing regimen respectively: 30, 45 and 60 mg/kg/day for 5 days a week. Results clearly show that in the absence of an adequate exposure to the chelating agent an appropriate clinical response cannot be achieved. The model provides also the opportunity to evaluate a priori the most suitable dosing regimen to achieve a desired therapeutic goal.

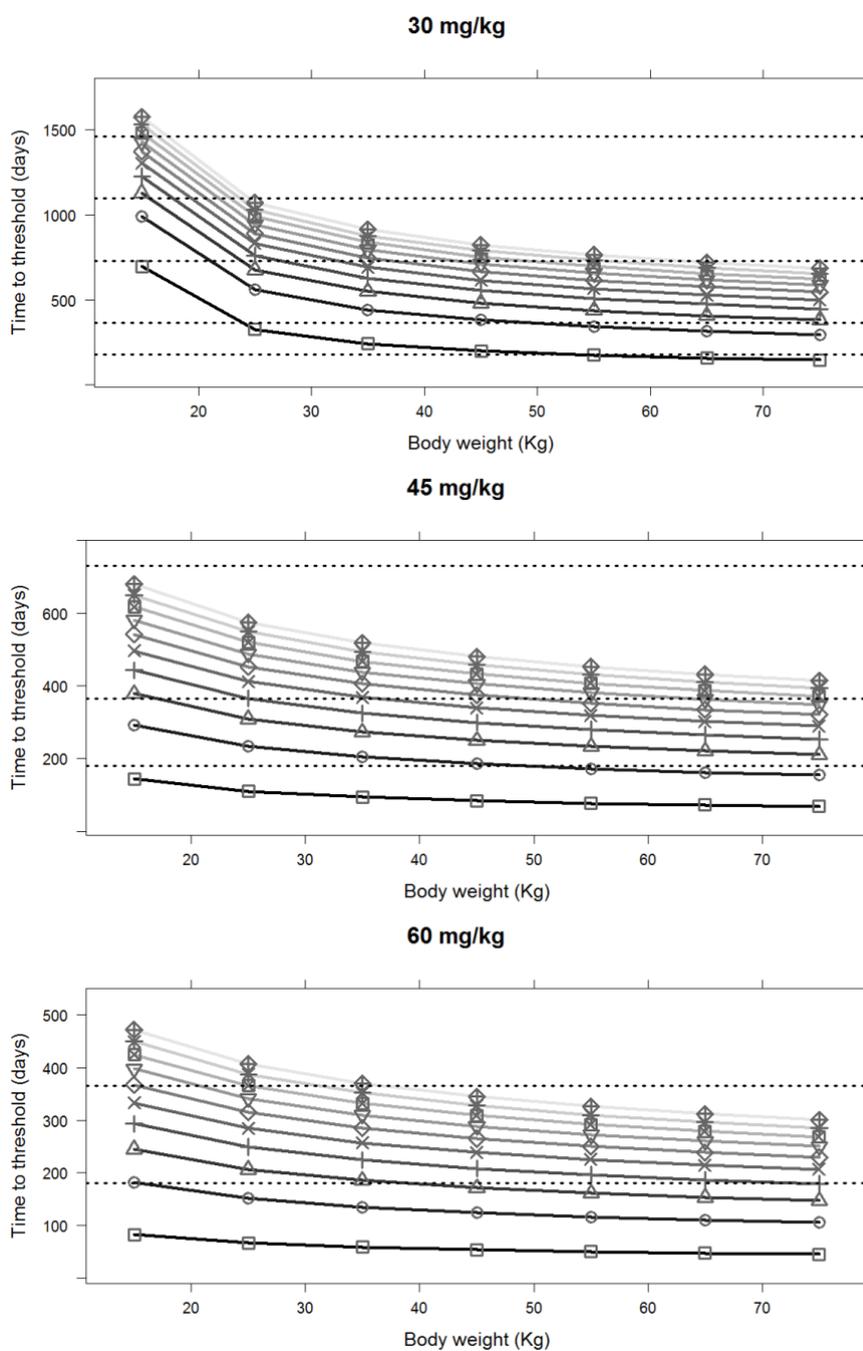


Figure 7. Time to reach a threshold of 2500 ug/L of serum ferritin based on different exposure levels in patients with different body weights (15 to 75 kg). The different panels show three scenarios where 30, 40 and 45 mg/kg dosing regimen have been evaluated. Each line represents a different starting baseline ferritin level (darker to lighter shows an increase in the starting baseline levels): square, circle, triangle with point up, plus, cross, diamond, triangle with point down, square cross, star and diamond plus represent respectively 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000 and 12000 ug/L of the starting baseline ferritin.

We have also investigated in one virtual patient of 45 kg receiving 45 mg/kg/day deferoxamine 5 days per week the impact of different compliance patterns in achieving a specific response, which was defined as time to reach the threshold of 2500 µg/L serum ferritin. Several conclusions can be derived from the results of the simulations: 1) if single doses are missed at random (reflecting poor quality of execution) (Figure 8 – scenario 1) as compared to doses missed consecutively (drug holidays) (Figure 8 – scenario 2) over a period of 1 year, a better and faster response is achieved; 2) if doses are missed consecutively over a given period of time, the shorter the period the better the clinical response as shown in Figure 8 – scenarios 2 to 5) in all the scenarios, if more than 60 % of the doses are missed (treatment compliance is lower than 40%) the therapeutic intervention is not effective; finally 4) a reduction in treatment compliance, especially when moving from 30 to 60% of missed doses clearly shows a significantly slower response indicating that even though the desired therapeutic outcome will be achieved the time to reach this goal might not be sustainable by the patient.

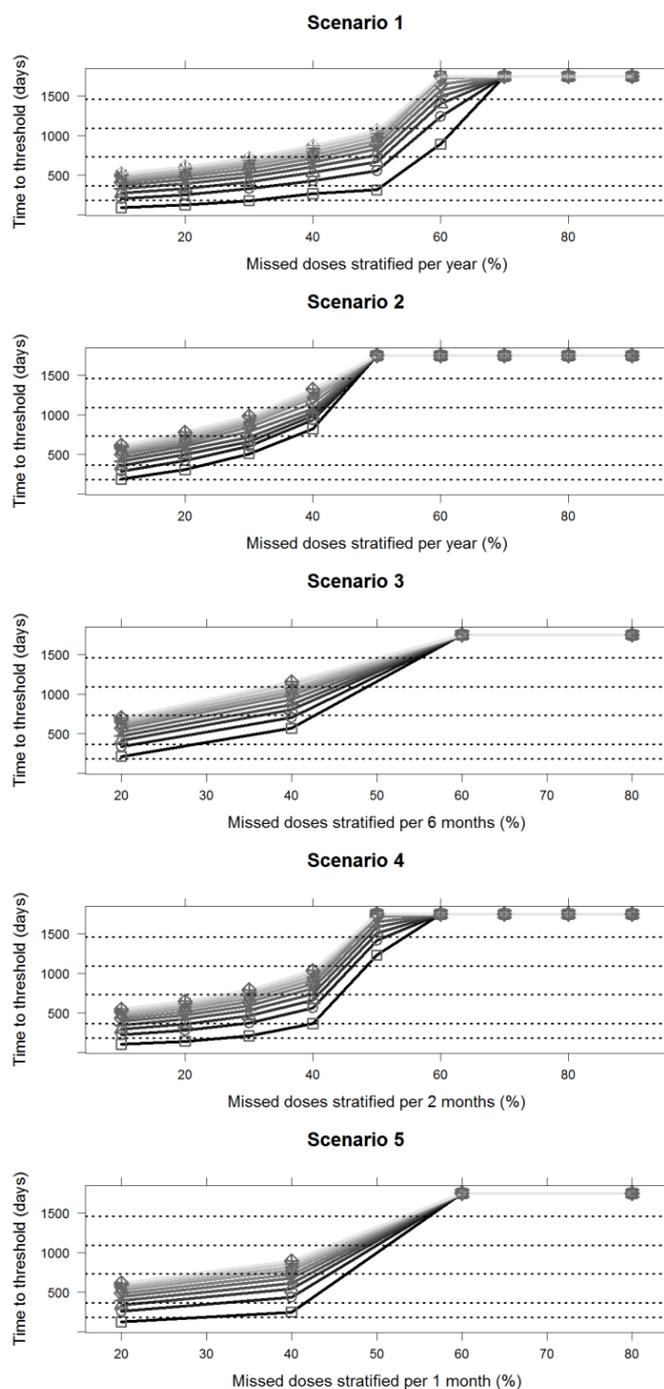


Figure 8. Time to reach a threshold of 2500 mcg/L of serum ferritin based on different compliance scenarios (10 to 90 % of missed doses). The different panels show five scenarios where different compliance patterns have been evaluated (see table II). Each line represents a different starting baseline ferritin level (darker to lighter shows an increase in the starting baseline levels): square, circle, triangle with point up, plus, cross, diamond, triangle with point down, square cross, star and diamond plus represent respectively 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000 and 12000 ug/L of the starting baseline ferritin.

7.4 Discussion

A model-based approach was proposed here to understand the implications of iron chelation therapy with deferoxamine on ferritin levels in patients affected by transfusion-dependent diseases. The complexity of the system requires an integrated approach that allows exploring the dynamics of the disease and its progression. A drug model was incorporated successfully into a disease model previously developed aimed at the characterisation of ferritin levels in this population. The model was evaluated using statistical and graphical criteria of goodness-of-fit and predictive performance measures as shown in table III and figures 3 to 6. The analysis reveals a strong effect of the disease status on the overall iron/ferritin conversion rate, and highlights also the role of treatment (drug exposure and compliance patterns) on the overall response and disease progression. In addition, the inclusion of IOV (57.4 %) allowed achieving a better description of individual profiles. Such a value of IOV appears to be influenced by larger intra-individual variability at higher serum ferritin levels where other mechanisms such as inflammatory disorders, and/or liver status play a role in determining the absolute ferritin value. Unfortunately, the lack of information on these variables allowed us only to take a stochastic approach for the quantification of such differences.

Parameterisation of iron chelation

Deferoxamine binds iron at different extracellular levels, and within the cell it targets lysosomal ferritin iron by stimulating ferritin degradation (23). Given that with the available data we could not distinguish among the different actions of deferoxamine, we decided to parameterise the drug effect as a proportional factor influencing the K_{out} of the turnover model. Furthermore, the same parameterisation would allow exploring the effect of other chelating agents: for example, oral chelators such as deferiprone and deferasirox also act intracellularly as deferoxamine, though targeting a different pathway (i.e., cytosolic ferritin iron) (23).

Clinical application

Model simulations were used to investigate the impact of different exposure levels and various compliance scenarios on serum ferritin levels. Results show that inadequate iron chelation therapy with sub-therapeutic exposure (Figure 7) as well as poor adherence to the assigned dosing regimen (Figure 8) would significantly increase the time required to achieve a desired clinical response, and in some cases (e.g., with treatment compliance lower than 40%) patients would not achieve at all the therapeutic goal. Even though these results might seem rather obvious, we are aware that clinically relevant changes in serum ferritin levels are observed over a long period of time and often crucial decisions have to be made before the clinical evidence is available. This availability of this model shows an opportunity to

explore different scenarios that have been so far evaluated empirically in clinical practice. For example, model simulations allow evaluating whether a compromise between lower exposure, aimed at a possible reduction in acute side effects is compensated by acceptable increase in the time to achieve the therapeutic goal. Likewise, it may be possible to evaluate the importance of different compliance patterns for the available chelating agents, yielding a more quantitative estimate of the changes in ferritin levels and /or risk of clinical failure. This information can then be used to support the decision making and to optimise the therapeutic intervention.

Limitations

Some limitations must be discussed in the context of this analysis. First of all, we used a PK model developed on literature data, which allowed us using mean population data and derive individual information based only on fixed allometric scaling. A more structured analysis of the PK of deferoxamine would reduce the uncertainty around the simulated exposure and would allow us explaining the variability in PK that propagates into the pharmacodynamics. On the other hand, we believe that the approach taken allowed us to characterise differences in the pharmacokinetics that we would have not been able to define only with information on the dosing regimen; e.g., changes in size are accounted for based on allometric scaling.

A second aspect is the role of compliance in the context of this analysis. The absence of quantitative data on treatment adherence in the population under investigation was a clear impediment. To overcome this issue we used the observed data to generate a variable that would allow us to have a gradient of treatment compliance. This decision was supported by the information available in the literature; we found clear evidence that high compliance leads to stable ferritin levels over time and that poor adherence to deferoxamine therapy is strongly correlated to a poor clinical outcome as nicely depicted in the work by Gabutti et al. (29) (Kaplan-Meier analysis presented in Figure 6). This was confirmed by a few other publications (20,30,31,36) and gave us the confidence that the approach taken would be robust enough to meet the purposes of this analysis.

Conclusion

In conclusion, we were able to gather further insights in the dynamics of a rather complex process such as iron overload using a model-based approach. Bearing in mind the limitations discussed and the relative level of uncertainty, the model has proven to be a useful tool to support decision making in clinical practice in the context of transfusion-dependent haemoglobinopathies. In addition, this approach will enable further evaluation of the dose rationale for existing and novel chelating agents.

References

1. Gibbons R, Higgs DR, Old JM, Olivieri NF, Swee Lay T, Wood WG. The Thalassaemia Syndromes - Fourth Edition. Blackwell Sci. 2001;
2. Galanello R, Origa R. Beta-thalassemia. *Orphanet J Rare Dis.* 2010 Jan;5:11.
3. Ginzburg Y, Rivella S. B-Thalassemia: a Model for Elucidating the Dynamic Regulation of Ineffective Erythropoiesis and Iron Metabolism. *Blood.* 2011 Oct 20;118(16):4321–30.
4. Rebullà P. Blood transfusion in beta thalassaemia major. *Transfus Med.* 1995 Dec;5(4):247–58.
5. Rebullà P, Modell B. Transfusion requirements and effects in patients with thalassaemia major. *Lancet.* 1991;337:277–80.
6. Rund D, Rachmilewitz E. Beta-Thalassemia. *N Engl J Med.* 2005;353:1135–46.
7. Thalassaemia International Federation (TIF). Guidelines for the clinical management of thalassaemia. 2008.
8. Porter J, Huehns E. Transfusion and exchange transfusion in sickle cell anaemias, with particular reference to iron metabolism. *Acta Haematol.* 1987;78:198–205.
9. Cunningham MJ, Macklin E a, Neufeld EJ, Cohen AR. Complications of beta-thalassemia major in North America. *Blood.* 2004 Jul 1;104(1):34–9.
10. Borgna-Pignatti C, Rugolotto S, De Stefano P, Zhao H, Cappellini MD, Del Vecchio G, et al. Survival and complications in patients with thalassaemia major treated with transfusion and deferoxamine. *Haematologica.* 2004;89(10):1187–93.
11. Borgna-Pignatti C, Cappellini MD, De Stefano P, Del Vecchio GC, Forni GL, Gamberini MR, et al. Survival and complications in thalassaemia. *Ann N Y Acad Sci.* 2005 Jan;1054:40–7.
12. Brittenham GM, Cohen a R, McLaren CE, Martin MB, Griffith PM, Nienhuis a W, et al. Hepatic iron stores and plasma ferritin concentration in patients with sickle cell anemia and thalassaemia major. *Am J Hematol.* 1993 Jan;42(1):81–5.
13. Lipschitz D, Cook J, Finch C. A clinical evaluation of serum ferritin as an index of iron stores. *N Engl J Med.* 1974;290(22):1213–6.
14. Puliyl M, Sposto R, Berdoukas V a, Hofstra TC, Nord A, Carson S, et al. Ferritin trends do not predict changes in total body iron in patients with transfusional iron overload. *Am J Hematol.* 2014 Apr;89(4):391–4.
15. Cappellini MD, Pattoneri P. Oral iron chelators. *Annu Rev Med.* 2009 Jan;60:25–38.

16. Kwiatkowski JL. Oral iron chelators. *Pediatr Clin North Am.* 2008 Apr;55(2):461–82.
17. Musallam KM, Taher AT. Iron chelation therapy for transfusional iron overload: a swift evolution. *Hemoglobin.* 2011 Jan;35(5-6):565–73.
18. Shander A, Sweeney J. Overview of Current Treatment Regimens in Iron Chelation Therapy. *US Hematol.* 2009;2(1):56–9.
19. Fisher S, Brunskill S, Doree C, Gooding S, Chowdhury O, Roberts D. Desferrioxamine mesylate for managing transfusional iron overload in people with transfusion-dependent thalassaemia (Review). *cochrane Libr.* 2013;(8).
20. Brittenham G, Griffith P, Nienhuis A, McLaren C, Young N, Tucker E, et al. Efficacy of deferoxamine in preventing complications of iron overload in patients with thalassemia major. *N Engl J Med.* 1994;331(9):567–73.
21. Novartis Pharmaceuticals UK. Deferoxamine summary of product characteristics.
22. Porter J. Deferoxamine pharmacokinetics. *Semin Hematol.* 2001 Jan;38(1 Suppl 1):63–8.
23. Theil EC. Mining ferritin iron: 2 pathways. *Blood.* 2009 Nov 12;114(20):4325–6.
24. Bentur Y, Koren G, Tesoro A, Carley H, Olivieri N, Freedman MH. Comparison of deferoxamine pharmacokinetics between asymptomatic thalassemic children and those exhibiting severe neurotoxicity. *Clin Pharmacol Ther.* 1990 Apr;47(4):478–82.
25. Porter JB, Rafique R, Srichairatanakool S, Davis B a, Shah FT, Hair T, et al. Recent insights into interactions of deferoxamine with cellular and plasma iron pools: Implications for clinical use. *Ann NY Acad Sci.* 2005 Jan;1054:155–68.
26. Comets E, Brendel K, Mentré F. Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. *Comput Methods Programs Biomed.* 2008 May;90(2):154–66.
27. Hooker AC, Staats CE, Karlsson MO. Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method. *Pharm Res.* 2007 Dec;24(12):2187–97.
28. Lindbom L, Ribbing J, Jonsson EN. Perl-speaks-NONMEM (PsN)--a Perl module for NONMEM related programming. *Comput Methods Programs Biomed.* 2004 Aug;75(2):85–94.
29. Gabutti V, Piga A. Results of Long-Term Iron-Chelating Therapy. *Acta Haematol.* 1996;95:26–36.
30. Galanello R, Kattamis A, Piga A, Fischer R, Leoni G, Ladis V, et al. A prospective randomized controlled trial on the safety and efficacy of alternating deferoxamine and deferiprone in the treatment of iron overload in patients with thalassemia. *Haematologica.* 2006;91:1241–3.

31. Olivieri N, Nathan D, MacMillan J, Wayne A, Liu P, McGee A, et al. Survival in Medically Treated Patients with Homozygous β -Thalassemia. *N Engl J Med.* 1994;331:574–8.
32. Modell B, Khan M, Darlison M. Survival in β -thalassaemia major in the UK: data from the UK Thalassaemia Register. *Lancet.* 2000;355(9220):2051–2.
33. Shampine L, Reichelt M. The MATLAB ODE Suite. *SIAM J Sci Comput.* 1997;18:1–22.
34. Shampine L, Reichelt M, Kierzenka J. Solving Index-1 DAEs in MATLAB and Simulink. *SIAM Rev.* 1999;41:538–52.
35. Piana C. Adherence to antiretroviral combination therapy in children. 2013. p. 181–92.
36. Kattamis A, Dinopoulos A, Ladis V, Berdousi H, Kattamis C. Variations of ferritin levels over a period of 15 years as a compliance chelation index in thalassemic patients. *Am J Hematol.* 2001 Dec;68(4):221–4.

CHAPTER 8

Model-based characterisation of the acute and long-term unfavourable effects of iron chelation therapy

F Bellanti, A Emerenciana, GC Del Vecchio, MC Putti, C Cosmi, I Fotzi, M Danhof, and O Della Pasqua

Ready for submission

Summary

Aims: The evaluation of the safety profile during the development of a drug is a challenging undertaking, especially as the drug-specific adverse events may be intertwined with disease-related complications. Using iron chelation therapy as an example, we propose a model-based approach to integrate epidemiological and pharmacological data for the characterisation of the acute, long-term adverse events and the disease complications due to transfusion-dependent iron overload.

Methods: Longitudinal data from a reference group of patients (n= 27) affected by β -thalassaemia major under chelation therapy with deferoxamine were evaluated in conjunction with literature data on the short-term safety profile of deferoxamine. Occurrence of the secondary co-morbidities hypothyroidism and diabetes mellitus was analysed based on a time to event approach in NONMEM v.7.2.0. In this analysis historical data were included as priors to reduce uncertainty in parameter estimates. Occurrence of the acute drug-specific adverse events arthralgia/myalgia and anaphylaxis were modelled as dose-dependent and dose-independent events.

Results: The predicted incidence for hypothyroidism and diabetes based on the hazard models with mean (90% CI) was 6.3% (0-14.8) and 8.9% (0-18.5), respectively. For a 45 mg/kg/day dose the mean (95% CI of the mean) simulated incidences for anaphylaxis and arthralgia/myalgia were 0.154% (0.139-0.169) and 21.01% (20.85-21.17) respectively; other doses as well as different compliance patterns were evaluated both for drug-specific AEs and disease complications.

Conclusions: A model-based approach provides the basis for a structured evaluation of the safety profile of drugs at different stages of development and for risk management, allowing integration of clinical and epidemiological data and consequently discrimination between the disease-related and the drug-related adverse events. Our simulations show that both chelation and transfusion history play a major role in determining the long-term adverse events and complications of disease. The findings also reveal a delicate balance between acute and long-term complications, indicating that inadequate chelation therapy or poor compliance can affect the desired therapeutic goal.

8.1 Introduction

In many chronic paediatric diseases such as transfusion-dependent haemoglobinopathies, where life-long red blood cell (RBC) transfusion is essential to survive (1–7), the direct and instantaneous therapeutic effectiveness needs to be balanced with long-term complications that depend both on the treatment intervention and the underlying disease progression. Two major aspects need to be considered when evaluating long-term effects: the disease progresses over time and may lead to disease specific complications and at the same time the frequency of drug-specific AEs may change over time or delayed, time-dependent AEs may be occur (8,9). Furthermore, in contrast to drug efficacy, even the short-term evaluation of drug-specific AEs can be extremely challenging, as data are often not available (e.g., a given event might not be observed during a clinical trial, especially if the incidence is relatively low) or not quantifiable due to recognised methodological issues (10–12). The two aspects very often overlap, making it rather difficult to discriminate the underlying cause. Lack of understanding of such an interaction may lead to inaccurate assessment of the safety profile of a drug. In fact, to fully characterise the safety profile, a variety of endpoints need to be considered in parallel, taking into account the correlations among them.

Chronic iron overload

Even though the management of the chronic RBC transfusion regimen and the availability of adequate iron chelation therapy have improved significantly in the last decades, patients with β -thalassemia will still experience a number of complications throughout their entire life (6,13,14).

Among the disease related complications, iron overload is the most clinically relevant and it is associated with several co-morbidities such as cardiac dysfunction, liver fibrosis, hypogonadism, hypothyroidism, hypoparathyroidism and diabetes mellitus (6,13,14). Cardiac disease caused by myocardial siderosis is the most relevant one, causing death in 71% of the patient population (15).

In the absence of an innate mechanism to remove the excess of iron, treatment with iron chelators is vital to prevent its accumulation and to manage the related complications (16–19). In addition to the disease related complications, the therapeutic intervention itself may also cause a number of undesired events (different for each iron chelator available on the market) that will play an essential role in the ability of the patient population not only to accept the intervention (poor adherence) but also to coexist with these complications for a life-long term. In this analysis we focus on the iron chelating agent deferoxamine (DFO), that is first-line therapy for transfusion-dependent diseases and has been available for the treatment of iron overload for more than 35 years (2,6,16–21). Among the various limitations of deferoxamine therapy recognised by clinicians and experts in the field

(6,17,19,22), compliance to the treatment plays a crucial role in the overall effectiveness of the treatment as well as the related complications.

Using DFO for the treatment of iron overload as an illustrative example, we propose and evaluate the advantages of a model-based approach for the characterisation of the safety profile of a medicinal product. We also show how modelling allows integration of epidemiological (literature) and pharmacological data for the quantification of the acute (drug specific) and long-term (disease specific) AEs of iron chelation therapy. Lastly, we show how the effect of treatment compliance can be assessed and correlated to acute and long-term events, disentangling the impact of inadequate chelation therapy from variable pattern of treatment compliance.

8.2 Methods

Data

To evaluate the model-based approach in the context of chronic iron overload we decided to select data in thalassaemic patients undergoing single therapy with deferoxamine and specifically collected data on incidence of hypothyroidism and diabetes mellitus. The choice of these two co-morbidities was made because they both are a clear consequence of the disease and no other influence of the drug therapy is expected except for the prevention of the complication itself. Furthermore, in the absence of clinical data on drug-specific AEs we simulated incidences for two extreme cases (i.e., arthralgia/myalgia as a very common AE and anaphylaxis as a rare AE) to assess their profiles after short- and long-term treatment. Specific details on the data are provided in the next few paragraphs.

Clinical Data on hypothyroidism and diabetes mellitus

The modelling analysis was performed using retrospective clinical data in 27 patients with β thalassaemia major from three different Italian centres: A.O. Universitaria Consorziale Policlinico di Bari U.O. Pediatria Federico Vecchio; A.O. Universitaria Policlinico di Sassari Clinica Pediatrica, ASL 1 D.H. per Talassemia; A.O. di Padova Clinica di Oncoematologia Pediatrica. The study has been conducted in full conformance with the principles of the Declaration of Helsinki and with the local laws and regulations concerning clinical trials. The protocol and the informed consent documents have been formally approved by the relevant research ethics committee of each clinical site.

Clinical data were collected retrospectively for a maximum of ten years in 27 patients affected by transfusion-dependent diseases, receiving deferoxamine as single drug for iron chelation therapy. Baseline characteristics of the patient population are provided in Table 1. Patients contributed with 40.2 observations on average (sd: 17), with a minimum of 4 samples per year.

Table 1. Baseline characteristics of the patient population (n=27)

	Units	Median	Range
Age	Years	14.6	6.8-19.9
Weight	Kg	46	17.5-71
Height	Cm	154	111-173
TSH	mIU/L	2.34	0.58-83.2
FT4	ng/dL	1.05	0.73-1.43
AST	U/L	33	7-159
ALT	U/L	56	9-372
Glucose	mg/dL	91	52-444
Creatinine	mg/dL	0.6	0.2-1.12
Ejection Fraction	%	64	35-77
Ferritin	µg/L	2260	393-8500

Literature data on co-morbidities and drug specific AEs (data abstraction)

A literature search has been performed to retrieve data on the incidence of hypothyroidism and diabetes mellitus in the thalassaemic population. At first reports from the Central Bureau of Statistics, The Netherlands were used as reference for the background incidence of the two co-morbidities in the overall population (23). Subsequently, a comprehensive literature search was performed using MESH terms in PubMed, in which articles describing hypothyroidism and diabetes in thalassaemic patients were retrieved. Thirteen articles in total (15,24–32,32–34) were identified with relevant information on the incidence of both co-morbidities. The keywords used comprised the names of the co-morbidity in combination with β -thalassaemia major, transfusional iron overload, deferoxamine, and a combination of them. In parallel, a separate search was performed on publications showing supporting evidence for the use of serum ferritin levels as a predictor for the occurrence of the two co-morbidities in thalassaemic patients (35–39). Of relevance is the finding that co-morbidity is with higher incidence in patients whose serum ferritin levels are consistently above 2500 µg/L (35–39). This threshold represents the boundaries for a shift in iron overload from moderate to a severe state (2,6,40–42). Given the level of detail provided by the authors, we have focused on the work by Belhoul et al. who demonstrated in a group of almost 400 patients a clear distinction in the incidence of hypothyroidism and diabetes mellitus in relation to a serum ferritin threshold of 2500 µg/L (38).

Finally for the evaluation of drug specific adverse events (anaphylaxis and arthralgia/myalgia), estimates reported on the summary of product characteristics (SPC) of DFO (20) were used to simulate the events of interest. The pharmacological classification proposed by Wills and Brown was used to select drug specific adverse events based on their frequency, time of onset and the relation with dose (43). Arthralgia/myalgia was selected as

an example of a very common Type A (dose-dependent) AE with a frequency greater than 1/10. In addition, anaphylaxis was identified as a rare, Type B (dose-independent) AE with a frequency between 1/10000 and 1/1000.

Modelling

Hazard models for hypothyroidism and diabetes mellitus

The models for hypothyroidism and diabetes were developed based on the combination of literature and clinical data. Three steps were taken for the development of each model:

- 1) An exponential hazard model was built based on literature data on the incidence of the co-morbidity in thalassaemic patients (disease effect) and in a healthy population (baseline);
- 2) The estimated parameters were used as priors to estimate the hazard in the retrospective clinical data in thalassaemic patients;
- 3) Literature data were used to incorporate the effect of serum ferritin levels as a covariate on the final hazard model.

During step 1, a time to event analysis was performed for both hypothyroidism and diabetes by implementing an exponential hazard model in NONMEM v.7.2 (Icon Development Solutions, USA). The initial model was based on the epidemiology reports and literature data and consisted in the following relationship between hazard and survival:

$$S(t) = e^{-\int_0^t h(t)dt} \quad \text{Equation 1}$$

Where the hazard is $h(t)$, and the survival (S) is a function of the cumulative hazard within the time interval 0 to t . The effect of disease was included as a covariate function (λ_{dis}) that would modify the background (initial) hazard (h_0) as follows:

$$h(t) = h_0(t) * e^{\lambda_{dis}} \quad \text{Equation 2}$$

In step 2 the normal-inverse Wishart prior (NWPRI) option was used in NONMEM (44) to estimate the incidence of the co-morbidities in the data collected during the retrospective study in thalassaemic patients. In the presence of extremely sparse data the use of prior information was deemed pivotal to ensure unbiased estimate of the disease effect and to stabilise the model.

Finally, in step 3 the work by Belhoul et al. (38) was used to justify the inclusion of serum ferritin levels as a covariate factor in the model developed in step 2. The objective was to

demonstrate that ferritin can be considered as a predictive factor for the probability (hazard) of developing co-morbidity. The threshold of 2500 $\mu\text{g/L}$ was used as reference value to dichotomise the data into two groups and ratio of the incidence of the co-morbidity in these groups was used to define the corresponding ratio in the hazard model. As shown in equation 2, the effect of the disease was described by two components depending on whether serum ferritin levels were above (λ_{frth}) or below (λ_{frtl}) the selected threshold:

$$h(t) = h_0(t) * e^{\lambda_{frth} + \lambda_{frtl}t} \quad \text{Equation 3}$$

A summary of the model building steps and parameter estimates for the hazard models of deferoxamine for hypothyroidism and diabetes mellitus is provided in Table 3.

Table 3. Model building steps and parameter estimates of the hazard model of deferoxamine for hypothyroidism and diabetes mellitus.

Hypothyroidism		
Parameter	Description	Estimate
Step 1: based on epidemiological and literature data		
h_0	Baseline hazard	0.000496
λ_{dis}	Disease as a predictor	2.69
Step 2: based on retrospective clinical data (step 1 used as prior)		
h_0	Baseline hazard	0.000496 (FIX)
λ_{dis}	Disease as a predictor	1.86
Step 3: based literature data (38)		
h_0	Baseline hazard	0.000496 (FIX)
λ_{frtl}	Disease when ferritin is below 2500 $\mu\text{g/L}$	1.03
λ_{frth}	Disease when ferritin is above 2500 $\mu\text{g/L}$	2.58
Diabetes mellitus		
Parameter	Description	Estimate
Step 1: based on epidemiological and literature data		
h_0	Baseline hazard	0.00036
λ_{dis}	Disease as a predictor	2.72
Step 2: based on retrospective clinical data (step 1 used as prior)		
h_0	Baseline hazard	0.00036 (FIX)
λ_{dis}	Disease as a predictor	2.54
Step 3: based literature data (38)		
h_0	Baseline hazard	0.00036 (FIX)
λ_{frtl}	Disease when ferritin is below 2500 $\mu\text{g/L}$	1.56
λ_{frth}	Disease when ferritin is above 2500 $\mu\text{g/L}$	3.33

Logistic models of acute drug specific adverse events

In contrast to the data fitting procedures used to describe the incidence of co-morbidities, drug-specific adverse events were evaluated by simulations using the information reported on the SPC of deferoxamine.

Two approaches were used to simulate the incidence of a very common dose-dependent AE (arthralgia/myalgia) and a rare dose-independent AE (anaphylaxis). In the first case, a logistic model with non-linear regression was developed correlating the drug levels a steady-state with the probability of adverse events in an exposure-dependent manner. Steady-state concentrations were simulated based on a PK model, which is described in later in this section. The logistic model was implemented as follows:

$$P = \frac{C_{ss}^{\gamma}}{(PC_{50}^{\gamma} + C_{ss}^{\gamma})} \quad \text{Equation 4}$$

where C_{ss} is the deferoxamine steady state concentration, PC_{50} is the concentration corresponding to a 50% probability of experiencing the AE, and γ is the coefficient defining the shape of the relationship. Parameter values for PC_{50} and γ were fixed to 13 $\mu\text{g}/\text{ml}$ and 2.5, respectively, to ensure that simulated incidence levels correspond to the figures reported in the SPC.

In the second case, a truncated normal distribution (with $x > 0$) was used in R to simulate a rare, dose-independent AE (anaphylaxis). The `rnorm` function (45) with mean equal to 0.5 and standard deviation equal to 0.5 was used to generate the probabilities of experiencing the adverse event. A severity of grade 2-3 was assumed for all AEs. However, for the purposes of this analysis no distinction was made between severity levels at the time of the event. Data was therefore summarised only as the overall frequency of AE.

Role of compliance

In a previous investigation we have highlighted the importance of treatment compliance for the effectiveness of drug therapy in patients with chronic iron overload [Chapter 7 of this thesis]. Poor adherence was found to have a major influence on the pharmacokinetics of the drug and subsequently on the desired clinical response. Compliance to treatment will therefore be one of the factors to be evaluated in the proposed simulation scenarios in order to assess its impact on the short- and long-term complications of iron chelation therapy.

Evaluation scenarios: clinical trial and not-in-trial Simulations

Simulations were performed to investigate the impact of different dose levels yielding to a range of exposure levels and various compliance scenarios on the onset and incidence of short and long-term unfavourable effects of iron chelation therapy (46). Data were

simulated for an overall period of maximum 10 years, where the 1st year is representative of a standard clinical trial and the subsequent years reflect a follow-up interval that has the objective of capturing real life conditions that may occur over a long-term period. Simulations were performed on a hypothetical patient population with similar demographic characteristics as the patients included in the retrospective clinical study (N=27) and were based on the models and final parameter estimates described above. Doses were adjusted according to changes in body weight at the scheduled visits. To ensure that uncertainty and variability in parameter estimates are accounted for, 250 simulations were performed for each individual in each of the scenarios described below. To facilitate visual representation of the simulated data, co-morbidity data were stratified by age in two major groups: above and below 12 years of age.

Simulation of drug concentrations and serum ferritin levels

A PK model and a PKPD model previously developed by our group [Chapter 7 of this thesis] were used to simulate deferoxamine exposure and serum ferritin levels. Deferoxamine concentrations (C_{ss}) were simulated based on a two compartment pharmacokinetic model with zero-order absorption and first-order elimination. Five dosing regimens (30, 40, 45, 50 and 60 mg/kg/day for 5 days a week) were evaluated and used as input for the logistic model (evaluation of short-term effects). These data were also used for the prediction of serum ferritin profiles, as described by the PKPD model. The predicted ferritin levels were incorporated as a covariate factor in the hazard models to evaluate the long-term complications of chelation therapy.

Clinical / Experimental conditions

To ensure the availability of clinically relevant scenarios, different deferoxamine dosing regimens yielding to a range of exposure levels were tested on patients starting at three different baselines ferritin levels, namely 1500, 2500 and 3500 $\mu\text{g/L}$ serum ferritin. This allowed further exploration of the correlation between ferritin levels and differences in compliance pattern. The three groups reflect well-defined populations of patients with poor chelation history (baseline at 3500 $\mu\text{g/L}$), patients with good chelation history (baseline at 1500 $\mu\text{g/L}$) and patients with unknown chelation history (baseline around 2500 $\mu\text{g/L}$).

In the initial set of simulations, exposure to deferoxamine was assumed to be constant over the course of treatment. In addition, it was assumed that all patients received the same dosing regimen: 45 mg/kg/day deferoxamine for 5 days a week. Treatment was maintained at constant dose levels for up to 10 years, under assumption of adequate or satisfactory response over time, even in those subjects showing initial ferritin levels above 3500 $\mu\text{g/L}$. Our main interest was to show how simulations can be used prospectively to evaluate long-term complications. Moreover, we demonstrate how these scenarios can be used to explore

the impact of variable treatment compliance. The selected scenarios are presented in Table 2.

Table 2. Simulation scenarios for the evaluation of different patterns of compliance.

		Number of missed doses in the stratification period				
		Single doses (Random)	Consecutive doses (Drug holidays)			
		Stratification				
	% of missed doses	1 year	1 year	6 months	2 months	1 month
Scenario 1	10%	25	25	/	5	/
Scenario 2	20%	50	50	25	10	5
Scenario 3	30%	75	75	/	15	/
Scenario 4	40%	100	100	50	20	10
Scenario 5	50%	125	125	/	25	/
Scenario 6	60%	150	150	75	30	15
Scenario 7	70%	175	175	/	35	/
Scenario 8	80%	200	200	100	40	20
Scenario 9	90%	225	225	/	45	/

Full adherence is equivalent to 250 doses per year

All the analyses described in the aforementioned paragraphs were performed in NONMEM version 7.2 (Icon Development Solutions, USA), with exception of the rare dose-independent AE, which was performed in R. All data manipulation, graphical and statistical summaries were performed in R (v.2.14.0).

8.3 Results

Hazard models for hypothyroidism and diabetes mellitus

Two survival models (exponential hazard) were developed for the quantification of hypothyroidism and diabetes in thalassaemic patients and used for prospective evaluation through model simulations. Figure 1 shows the predictions for hypothyroidism and diabetes, as compared to the available epidemiological and literature data, as described previously in step 1. Both models show good agreement with the observed data.

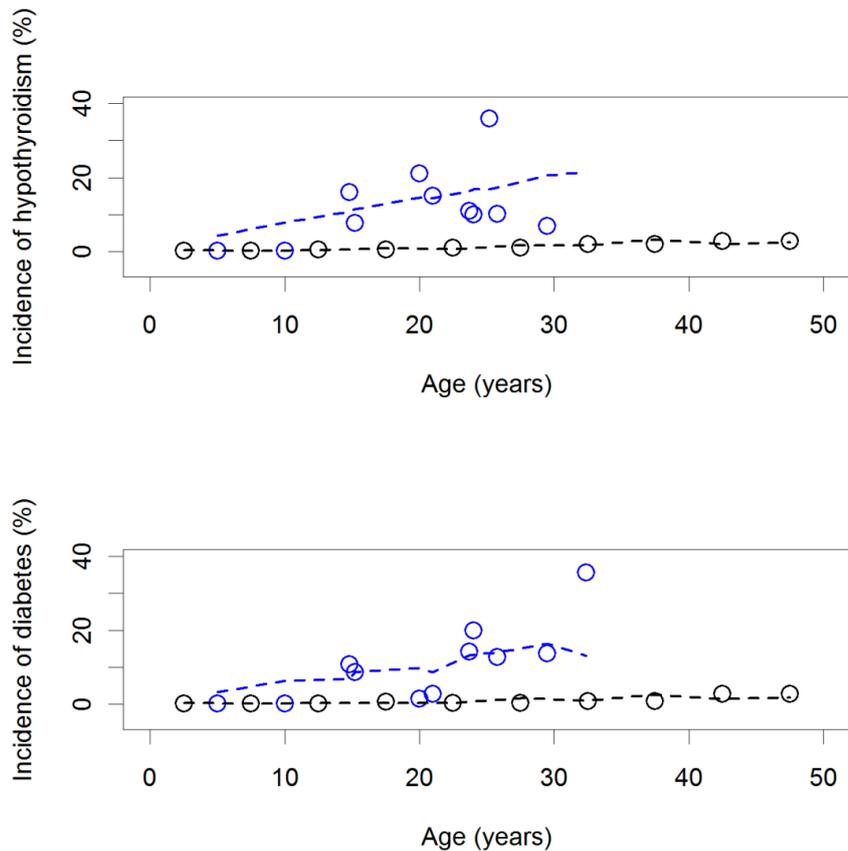


Figure 1. Performance of the hazard model for hypothyroidism (top panel) and diabetes (bottom panel) based on modelling of historical data. Black circles represent observed literature data for baseline incidence of the co-morbidities in the overall population whereas blue circles represent the observed literature incidence for thalassaemic patients. The dashed lines show model predictions in blue with respect to the patient population and in black with respect to the baseline incidence.

The results of the final model are presented in Figure 2, (after step 3: inclusion of serum ferritin as a predictor of the co-morbidity), in which published literature (15,24,32,38,39,47) data was included as prior for the analysis of the available clinical data. The mean (90% CI) predicted incidence of hypothyroidism and diabetes was 6.3% (0-14.8) and 8.9% (0-18.5), respectively. Both models were considered adequate for simulation purposes.

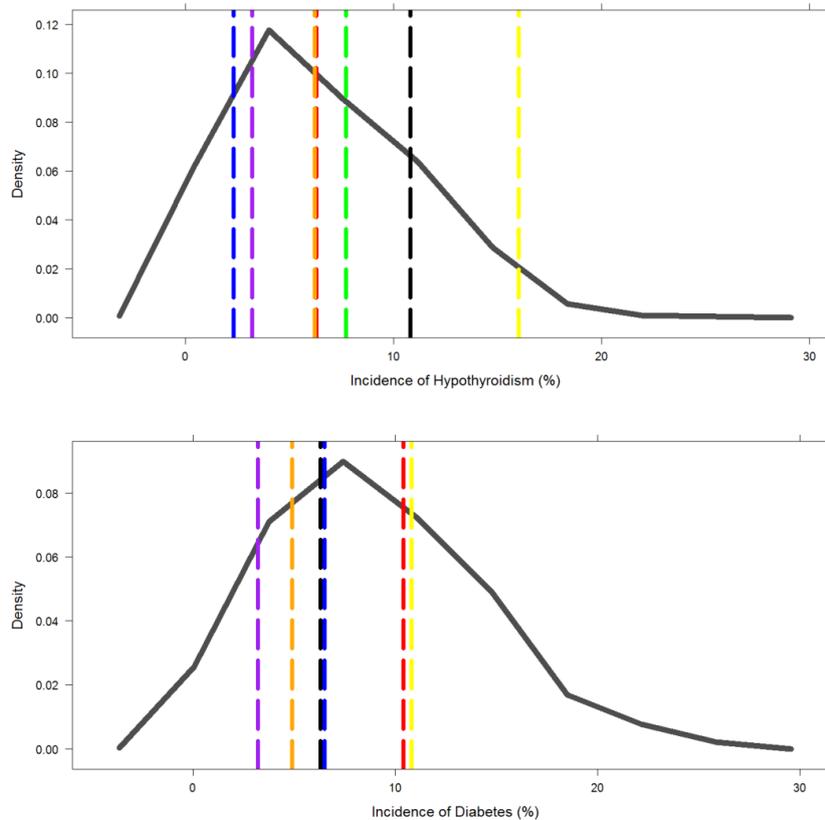


Figure 2. Validation of the hazard models for hypothyroidism (top panel) and diabetes (bottom panel). Model predicted incidence (solid dark grey line) is compared to literature data (coloured dashed lines): Borgna-Pignatti et al (black); Belhoul et al (red); Mehrvar et al (blue); Aydinoc et al (yellow); Shamshirsaz et al (green); and Kyriakou et al (orange and purple).

Clinical trial and not-in-trial simulations

The results of the evaluation of drug and treatment compliance levels on long term disease progression are presented in Figure 3. Simulation of the incidence of hypothyroidism and diabetes in a virtual population of 27 patients are stratified by age groups (below or above 12 years of age at start of treatment). In patients below 12 years of age a slight negative trend is observed indicating a reduction in the incidence of the co-morbidities with increasing dose levels; this is not the case in the other group where no significant changes are observed. Furthermore, a clear distinction in the incidence of both co-morbidities was observed for patients with different starting baseline ferritin levels. This finding highlights the relevance of transfusion and chelation history for these outcomes.

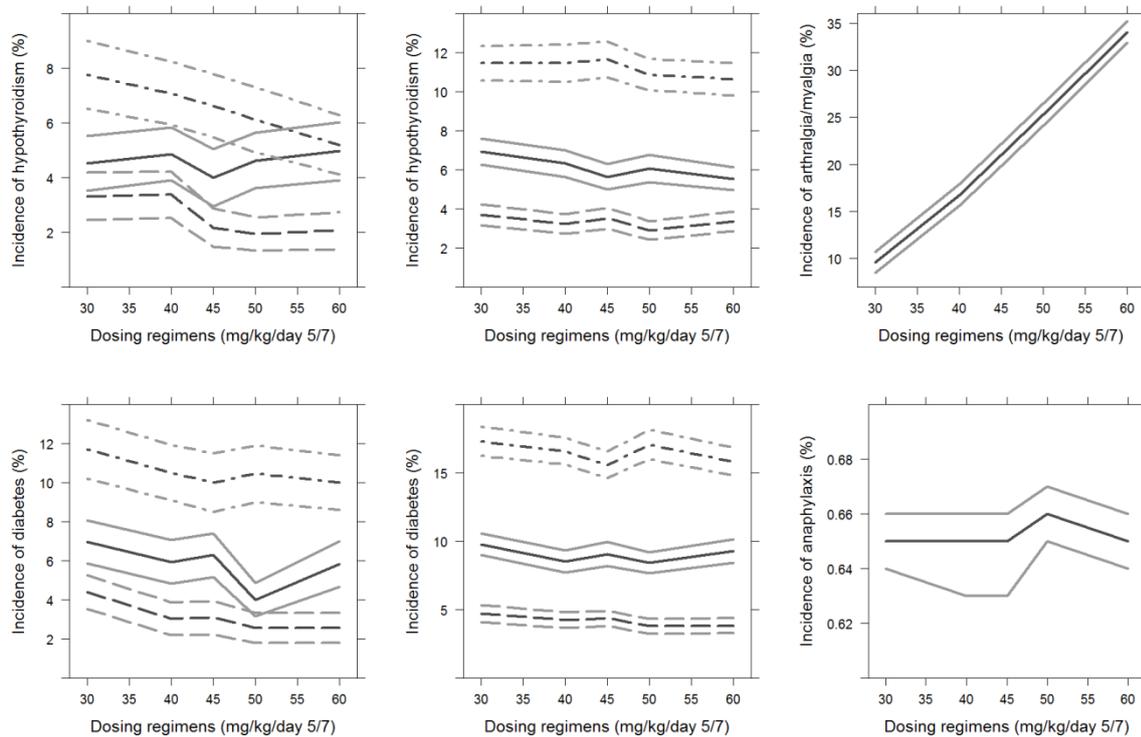


Figure 3. Effect of different exposure levels (x axes show different dosing regimens investigated in the simulation) on hypothyroidism, diabetes, arthralgia/myalgia and anaphylaxis in the population under investigation for an observational period of maximum 10 years. *Top left and top mid panels* show the simulations outcome for hypothyroidism after stratification of the patients into two age categories, i.e., below and above 12 years of age, respectively. *Bottom left and bottom mid panels* show the simulations outcome for diabetes in patients below and above 12 years of age, respectively. The dashed, solid and dotted-dashed lines represent respectively the three subgroups of patients with adequate, unknown and poor chelation history. The top right panel show the results for arthralgia/myalgia, whereas the bottom right panel gives the results for anaphylaxis. In all scenarios the dark grey lines represent the mean and the light grey lines represent the 95% confidence interval of the mean.

On the other hand, the simulations describing the occurrence of acute, drug-specific AEs the show the implications of dose-dependent and dose-independent adverse events, on the individual safety profile of each patient, with the incidence of myalgia/arthralgia increasing proportionally with the dose of deferoxamine.

Given the interaction between treatment response, as determined by ferritin levels and adherence to the prescribed dosing regimen, we also included an evaluation of the impact of different compliance patterns. Results are shown in Figures 4, 5 and 6 for hypothyroidism, diabetes and arthralgia/myalgia and anaphylaxis, respectively. Similarly to what we have observed when evaluating the implications of different exposure levels, stratification of the

data by age indicates an increase in the incidence of hypothyroidism and diabetes with decreasing levels of adherence in patients below 12 years of age (Figures 4 and 5 – left panels). Similar trends are observed among the different scenarios proposed. In addition, stratification based on starting ferritin levels shows the importance of the patient's treatment history for the prediction of long-term complications.

When looking at arthralgia/myalgia (Figure 6 – left panels) the different scenarios are characterised by similar profiles, i.e., with increasing incidence of adverse events at increasing doses; but the magnitude of the effect is slightly altered at different levels of compliance. By contrast, no major differences are observed for the dose-independent AE (anaphylaxis: Figure 6 – right panels).

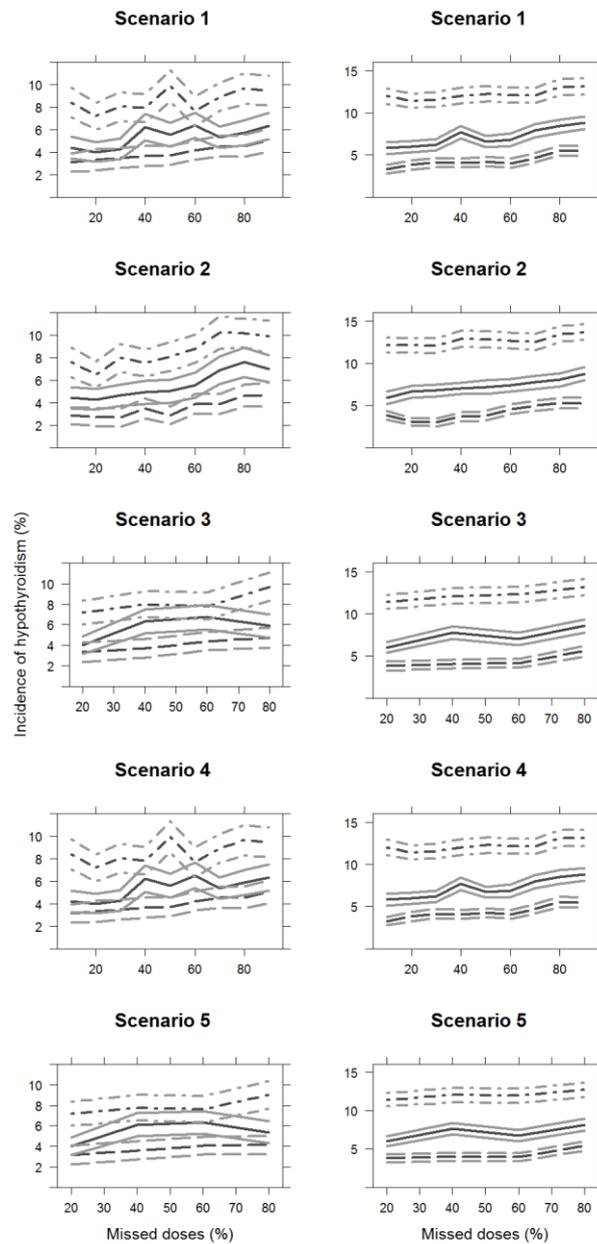


Figure 4. Effect of different patterns of compliance on hypothyroidism in the population under investigation for an observational period of maximum 10 years. Left and right panels show results based on stratification of patients into two age categories, i.e., below and above 12 years of age, respectively. In all scenarios: the dashed, solid and dotted-dashed lines represent respectively the three subgroups of patients with adequate, unknown and poor chelation history; the dark grey lines represent the mean and the light grey lines represent the 95% confidence interval of the mean. The 5 scenarios presented here are detailed in the methods and in Table 2.

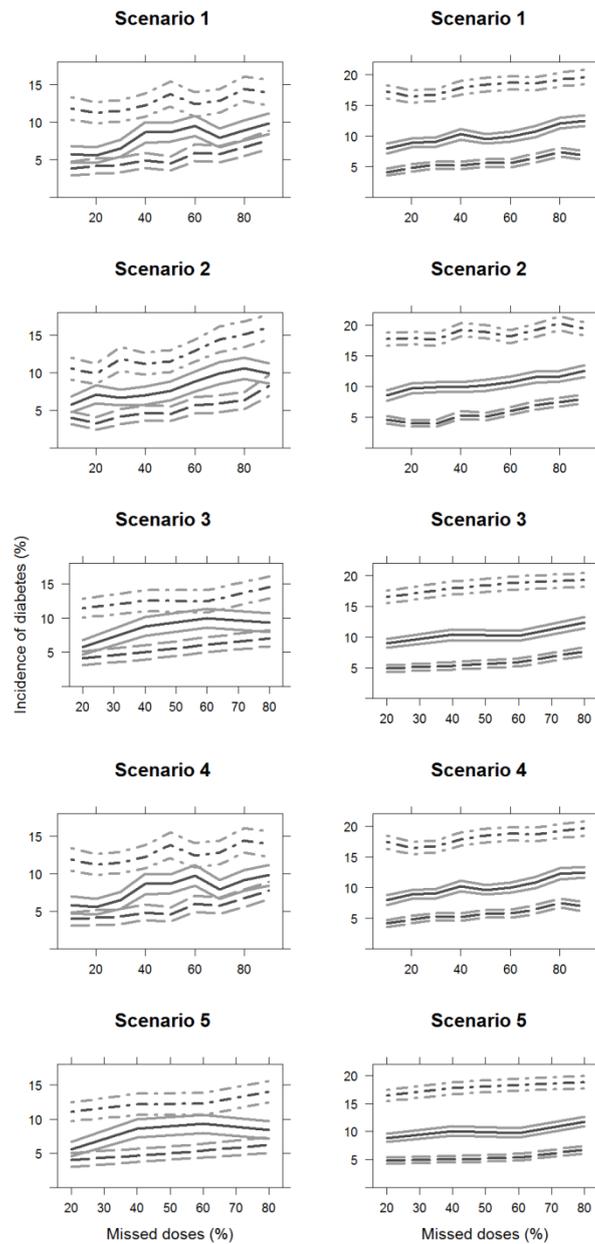


Figure 5. Effect of different patterns of compliance on diabetes in the population under investigation for an observational period of maximum 10 years. Left and right panels show results based on patient stratification into two age categories, i.e., below and above 12 years of age, respectively. In all scenarios: the dashed, solid and dotted-dashed lines represent the three subgroups of patients with adequate, unknown and poor chelation history; the dark grey lines represent the mean and the light grey lines represent the 95% confidence interval of the mean. The 5 scenarios presented here are detailed in the methods and in Table 2.

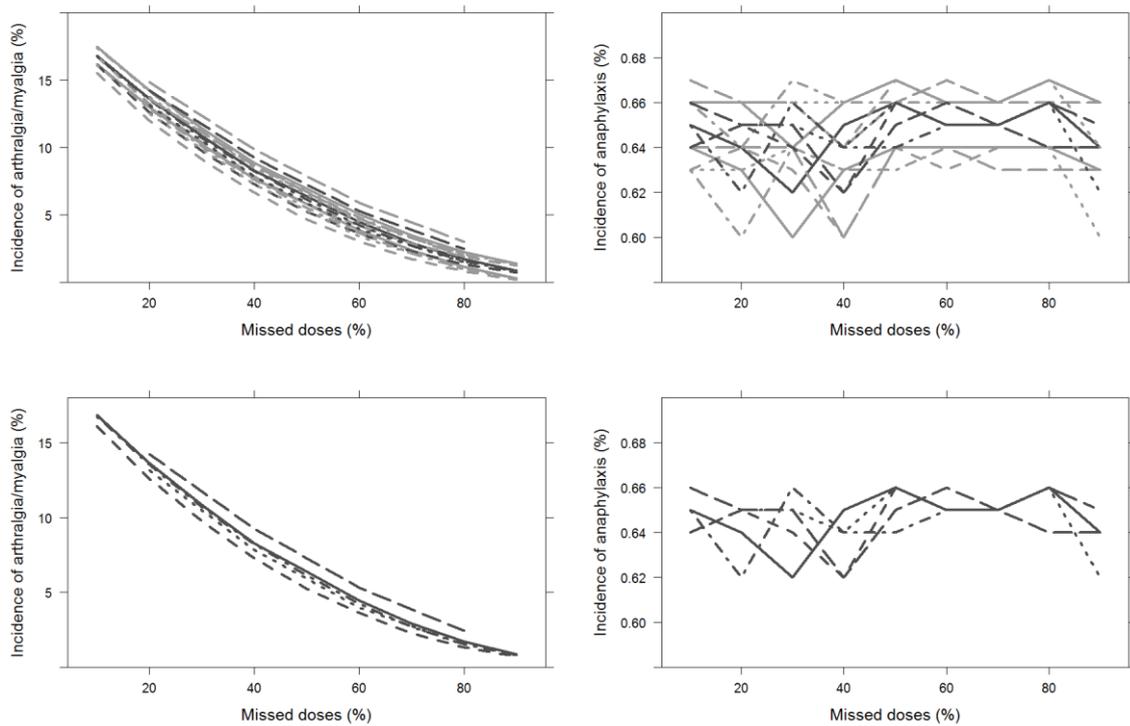


Figure 6. Effect of different compliance patterns on arthralgia/myalgia (left panels) and anaphylaxis (right panels) in the population under investigation for an observational period of maximum 10 years. The solid, dashed (small), dotted, dotted-dashed, and dashed (large) lines represent the scenarios investigated from 1 to 5 respectively. In all panels the dark grey lines represent the mean and the light grey lines represent the 95% confidence interval of the mean. Mean and 95%-CI of the mean are presented in the top panels, whereas the bottom panels show only the mean value.

8.4 Discussion and Conclusion

A model-based approach was implemented to evaluate simultaneously the short- and long-term unfavourable effects of iron chelation therapy. Epidemiological and pharmacological data have been combined to appropriately estimate the parameters of interest in the survival models. In contrast to traditional data meta-analysis, summary statistics is used to integrate data from different sources; here we rely on literature summaries to fit a population model for events whose incidence is relatively low to be derived from individual clinical trials. In fact, the epidemiological data was deemed essential for an unbiased quantification of the incidence of hypothyroidism and diabetes in the thalassaemic population. Both models were successfully validated, as shown in Figures 1 and 2.

Whereas external model validation procedures could not be easily implemented for this type of analysis, basic diagnostics plots suggested that the model was sufficiently robust support its use for simulation purposes. In fact, the use of literature summaries, i.e., point estimates,

as reference input data for fitting has been applied previously in a number of therapeutic areas in investigations with similar scope (48–50).

Treatment as a disease modifying factor

Our simulations show that long-term complications associated with inadequate chelation due to suboptimal dosing or poor compliance has major implications for patients in the lower age group, whereas almost no effect is observed for those patients in the higher age group. These results suggest that this phenomenon is partially masked by the baseline age of the population, which in turn reflects the chelation history of the patients. This is also evident by the difference in the overall incidence of the comorbidities among the three subgroups evaluated in each scenario, depending on their chelation history. These findings are in agreement with previous report on the consequences and cost of noncompliance to iron chelation (51,52). In clinical practice, improvement of compliance with chelation therapy is considered the best prevention for hypothyroidism. Guidelines also recommend regular follow-up and optimising chelation therapy in patients showing sub-clinical hypothyroidism, i.e., basal levels of TSH 5 to 7 mUI/ml.

In theory, our analysis suggests that changes in the treatment of patient at a late phase of the disease could potentially have little or no impact on the probability of developing hypothyroidism or diabetes. Hence, effective treatment at the start of chelation therapy may determine long-term onset of co-morbidities. Whilst the proposed simulations scenarios have been limited to a predefined set of compliance patterns with overall dose intake ranging from 10% (worst case scenario) to 90 % in patients with perfect adherence to treatment, literature data on deferoxamine reveals that mean compliance in patients ranges from 59 to 78 % (51).

The proposed scenarios also provide an opportunity to assess prospectively the correlation between short- and long-term complications. For instance, until now it is unclear whether changes in the dosing regimen can be implemented to provide benefit for a given patient in the short-term without significantly affecting the long-term disease progression. Such a correlation can be seen in the scenarios shown in Figures 3 to 5 for the long-term complications. Focus of treatment is mostly on correcting for changes in ferritin levels, but dose rationale currently does not assess how different dose levels may lead to higher or lower incidence of long term co-morbidities.

Limitations

The simulation scenarios presented here represent a simplification of a complex therapeutic reality in which the nature and number of co-morbidities and drug-specific AEs are much higher than those included in our analysis (53–55). Nevertheless, we believe that the

selection of a subset of AEs has enabled us to demonstrate how inferences by modelling and simulation can be used to characterise the overall safety profile of a compound. Furthermore, our approach shows how to explore safety concerns pro-actively in a quantitative manner even in the absence of sufficient data from randomised clinical trials. We acknowledge, however, that the lack of available clinical data imposed the integration of epidemiological and literature data to develop the final models based on population summary data, which may mask some specific features of the disease or treatment at the individual patient level, especially if one takes into account potential correlations or interaction between covariates. Therefore, the impact of such an interaction, as well as of the correlation between endpoints could not be evaluated. The availability of more informative, individual patient data could have provided further support for our assumptions, but we do not anticipate that such data would alter the final conclusions from the proposed simulation scenarios.

The shortcoming from individual data may have been compensated by the incorporation of time-dependent effects (and covariate factors), which allowed a clear distinction between disease-specific (long term) and drug-specific (short term) AEs.

A possible weakness remains in that very few data were available from long term safety follow up studies including paediatric and adults. Such data might have provided better estimates of the parameters and covariate factors determining the timing and age of onset of co-morbidities.

We also acknowledge that the stratification of AEs by their grade of severity would be more relevant in clinical practice. Here we have assumed a grade 2-3 for all simulated AEs to reduce the complexity of the scenarios. The same applies for the duration of the AEs and the clinical implications that the event would have for individual patients, such as dose titration over even change of chelator. This simplification was necessary to ensure that focus were given the time-dependencies associated with the long term consequences of inadequate chelation therapy (56,57).

Perspectives

In this analysis we showed that M&S provides the necessary tools to overcome the methodological and practical hurdles in the evaluation of the safety profile of a compound. We foresee the advantages of applying such an approach in the context of a full benefit-risk (BR) appraisal, where the lack of a systematic and more structured approach is acknowledged by different parties (58–62). Of particular relevance for the implementation of BR assessment, is the possibility of exploring rare dose-independent AEs. It is worth mentioning that in controlled trials and especially in paediatric trials, where limited numbers of patients are enrolled, these events might not even be observed. We believe that in such cases, modelling & simulation enables the integration of available information (e.g.,

extrapolation of adult data) to explore in a quantitative manner the implication of (clinically relevant) what-if scenarios (10,11).

Despite the limited number of scenarios presented here, several aspects can be considered and analysed by clinical trial and not-in-trial simulations. Such a framework may allow common questions in paediatric research to be evaluated in a systematic way, especially those related to developmental growth or age, which may lead to changes in the incidence of AEs over time. Another important application is the assessment of susceptibility of subgroups or population minorities which may not be appropriately represented in the trial population (63).

Conclusions

In summary, our investigation has illustrated the advantages of a model-based approach for the characterisation of the safety profile of drug in children. The use of modelling and simulation does not only provide the basis for the systematic integration of clinical and epidemiological data as a means to overcome the limited data availability in this population, but also allows one to disentangle disease-specific from the drug-specific adverse events, which are often intertwined, but have different impact on long-term outcome of treatment. Irrespective of the level of understanding or the mechanisms underlying adverse events, the availability of a simulation framework to evaluate the safety profile of a treatment offers a unique opportunity to explore scenarios which may not be feasible or even acceptable in real life, but which nevertheless provide insight into the role of the drug, the patient and the disease in the outcome of an intervention. Such information may be essential for accurate assessment of the benefit-risk profile of a medicinal product in children.

References

1. Gibbons R, Higgs DR, Old JM, Olivieri NF, Swee Lay T, Wood WG. *The Thalassemia Syndromes - Fourth Edition*. Blackwell Sci. 2001;
2. Galanello R, Origa R. Beta-thalassemia. *Orphanet J Rare Dis*. 2010 Jan;5:11.
3. Ginzburg Y, Rivella S. B-Thalassemia: a Model for Elucidating the Dynamic Regulation of Ineffective Erythropoiesis and Iron Metabolism. *Blood*. 2011 Oct 20;118(16):4321–30.
4. Rebullà P. Blood transfusion in beta thalassaemia major. *Transfus Med*. 1995 Dec;5(4):247–58.
5. Rebullà P, Modell B. Transfusion requirements and effects in patients with thalassaemia major. *Lancet*. 1991;337:277–80.
6. Rund D, Rachmilewitz E. Beta-Thalassemia. *N Engl J Med*. 2005;353:1135–46.
7. TIF. Guidelines for the clinical management of thalassaemia. 2008.
8. Edwards I, Aronson J. Adverse drug reactions: definitions, diagnosis, and management. *Lancet*. 2000;356(9237):1255–9.
9. Aronson J, Ferner R. Joining the DoTS: new approach to classifying adverse drug reactions. *BMJ*. 2003;327(7425):1222–5.
10. Fescharek R, Nicolay U, Arras-Reiter C. Monitoring and safety assessment in Phase I to III clinical trials. *Dev Biol Stand*. 1998;95:203–9.
11. Wahab I, Pratt N, Kalisch L, Roughead E. The detection of adverse events in randomized clinical trials: can we really say new medicines are safe? *Curr Durg Saf*. 2013;8(2):104–13.
12. Sheiner L. Learning versus confirming in clinical drug development. *Clin Pharmacol Ther*. 1997;61(3):275–91.
13. Cunningham MJ, Macklin E a, Neufeld EJ, Cohen AR. Complications of beta-thalassemia major in North America. *Blood*. 2004 Jul 1;104(1):34–9.
14. Borgna-Pignatti C, Rugolotto S, De Stefano P, Zhao H, Cappellini MD, Del Vecchio G, et al. Survival and complications in patients with thalassemia major treated with transfusion and deferoxamine. *Haematologica*. 2004;89(10):1187–93.
15. Borgna-Pignatti C, Cappellini MD, De Stefano P, Del Vecchio GC, Forni GL, Gamberini MR, et al. Survival and complications in thalassemia. *Ann N Y Acad Sci*. 2005 Jan;1054:40–7.
16. Cappellini MD, Pattoneri P. Oral iron chelators. *Annu Rev Med*. 2009 Jan;60:25–38.

17. Kwiatkowski JL. Oral iron chelators. *Pediatr Clin North Am.* 2008 Apr;55(2):461–82.
18. Musallam KM, Taher AT. Iron chelation therapy for transfusional iron overload: a swift evolution. *Hemoglobin.* 2011 Jan;35(5-6):565–73.
19. Shander A, Sweeney J. Overview of Current Treatment Regimens in Iron Chelation Therapy. *US Hematol.* 2009;2(1):56–9.
20. Novartis Pharmaceuticals UK. Deferoxamine summary of product characteristics.
21. Theil EC. Mining ferritin iron: 2 pathways. *Blood.* 2009 Nov 12;114(20):4325–6.
22. Bentur Y, Koren G, Tesoro A, Carley H, Olivieri N, Freedman MH. Comparison of deferoxamine pharmacokinetics between asymptomatic thalassemic children and those exhibiting severe neurotoxicity. *Clin Pharmacol Ther.* 1990 Apr;47(4):478–82.
23. CBS. Centraal Bureau voor de Statistiek. Available from: <http://www.cbs.nl/nl-NL/menu/home/default.htm>
24. Aydinok Y, Darcan S, Polat A, Kavakli K, Nisli G, Coker M, et al. Endocrine Complications in Patients with Beta-thalassemia. *J Trop Pediatr.* 2002;48(February).
25. Gulati R, Bhatia V, Agarwal SS. Early Onset of Endocrine Abnormalities in β -Thalassemia Major in a Developing Country. *J Pediatr Endocrinol Metab.* 2000 Jan;13(6):651–6.
26. Farmaki K. Hypothyroidism in Thalassemia. In: Springer D, editor. *Hypothyroidism - Influences and Treatments.* 2012. p. 97–110.
27. Ghader F, Kousarian M, Farzin D. High-dose deferoxamine treatment (intravenous) for thalassaemia patients with cardiac complications. *East Mediterr Heal J.* 2007;13(5):1053–9.
28. Fung EB, Harmatz PR, Lee PDK, Milet M, Bellevue R, Jeng MR, et al. Increased prevalence of iron-overload associated endocrinopathy in thalassaemia versus sickle-cell disease. *Br J Haematol.* 2006 Nov;135(4):574–82.
29. Lai ME, Grady RW, Vacquer S, Pepe A, Carta MP, Bina P, et al. Increased survival and reversion of iron-induced cardiac disease in patients with thalassemia major receiving intensive combined chelation therapy as compared to desferoxamine alone. *Blood Cells Mol Dis.* Elsevier Inc.; 2010 Aug 15;45(2):136–9.
30. Irshaid F, Mansi K. Status of Thyroid Function and Iron Overload in Adolescents and Young Adults with Beta- Thalassemia Major Treated with Deferoxamine in. *Int J Biol life Sci.* 2011;7(1):47–52.
31. 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 Mar;28(2):196–208.

32. Shamshirsaz AA, Bekheirnia MR, Kamgar M, Pourzahedgilani N, Bouzari N, Habibzadeh M, et al. Metabolic and endocrinologic complications in beta-thalassemia major: a multicenter study in Tehran. *BMC Endocr Disord.* 2003 Aug 12;3(1):4.
33. Fung EB, Xu Y, Kwiatkowski JL, Vogiatzi MG, Neufeld E, Olivieri N, et al. Relationship between chronic transfusion therapy and body composition in subjects with thalassemia. *J Pediatr.* Mosby, Inc.; 2010 Oct;157(4):641–7, 647.e1–2.
34. Swaminathan S, Fonseca VA, Alam MG, Shah S V. The Role of Iron in Diabetes and Its Complications. *Diabetes Care.* 2007;30:1926–33.
35. Dmochowski K, Finegood DT, Francombe W, Tyler B, Zinman B. Factors determining glucose tolerance in patients with thalassemia major. *J Clin Endocrinol Metab.* 1993;77(2):478–83.
36. Li M, Peng SS, Lu M, Chang H, Yang Y, Jou S, et al. Diabetes Mellitus in Patients With Thalassemia Major. *Pediatr blood cancer.* 2014;61:20–4.
37. Chern JPS, Lin KL, Lu M, Lin D, Lin K, Chen J, et al. Abnormal Glucose Tolerance in Transfusion-Dependent Beta-Thalassemic Patients. *Diabetes Care.* 2001;24:850–4.
38. Belhouel KM, Bakir ML, Saned M-S, Kadhim AM a, Musallam KM, Taher AT. Serum ferritin levels and endocrinopathy in medically treated patients with β thalassemia major. *Ann Hematol.* 2012 Jul;91(7):1107–14.
39. Mehrvar a, Azarkeivan a, Faranoush M, Mehrvar N, Saberinedjad J, Ghorbani R, et al. Endocrinopathies in patients with transfusion-dependent beta-thalassemia. *Pediatr Hematol Oncol.* 2008;25(3):187–94.
40. Brittenham G, Griffith P, Nienhuis A, McLaren C, Young N, Tucker E, et al. Efficacy of deferoxamine in preventing complications of iron overload in patients with thalassemia major. *N Engl J Med.* 1994;331(9):567–73.
41. Modell B, Khan M, Darlison M. Survival in β -thalassaemia major in the UK: data from the UK Thalassaemia Register. *Lancet.* 2000;355(9220):2051–2.
42. Olivieri N, Nathan D, MacMillan J, Wayne A, Liu P, McGee A, et al. Survival in Medically Treated Patients with Homozygous β -Thalassemia. *N Engl J Med.* 1994;331:574–8.
43. Wills S, Brown D. A proposed new means of classifying adverse drug reactions to medicines. *Pharm J.* 1999;262:163–5.
44. Boeckmann AJ, Sheiner LB, Beal SL. *NONMEM Users Guide - Part VIII.* 2011.
45. R Core Team. The R stats package - The log normal distribution - <https://stat.ethz.ch/R-manual/R-patched/library/stats/html/Lognormal.html>.

46. Chain ASY, Dieleman JP, van Noord, Charlotte Hofman A, Stricker BHC, Danhof M, Sturkenboom MCJM, et al. Not-in-trial simulation I: Bridging cardiovascular risk from clinical trials to real-life conditions. *Br J Clin Pharmacol*. 2013;76(6):964–72.
47. Kyriakou A, Skordis N. Thalassaemia and Aberrations of Growth and Puberty. *Mediterr J Hematol Infect Dis*. 2009;1(1).
48. Stroh M, Green M, Cha E, Zhang N, Wada R, Jin J. Meta-analysis of Published Efficacy and Safety Data for Docetaxel in Second-Line Treatment of Patients with Advanced Non-Small-Cell Lung Cancer. PAGE meeting. 2014.
49. Maringwa J, Cox E, Harnisch L, Gao X. Model-based meta-analysis of summary clinical outcome data in idiopathic pulmonary fibrosis (IPF). PAGE meeting. 2012.
50. Kathman S, Williams D, Hodge J, Dar M. A Bayesian population PK-PD model for ispinesib/docetaxel combination-induced myelosuppression. *Cancer Chemother Pharmacol*. 2009;63(3):469–76.
51. Delea T, Edelsberg J, Sofrygin O, Thomas S, Baladi J, Phatak P, et al. Consequences and costs of noncompliance with iron chelation therapy in patients with transfusion-dependent thalassemia: a literature review. *Transfusion*. 2007;47(10):1919–29.
52. De Sanctis V, Soliman A, Elsedfy H, Skordis N, Kattamis C, Angastiniotis M, et al. Growth and endocrine disorders in thalassemia: The international network on endocrine complications in thalassemia (I-CET) position statement and guidelines. *Indian J Endocrinol Metab*. 2013;17(1):8–18.
53. Farmaki K, Tzoumari I, Pappa C. Oral chelators in transfusion-dependent thalassemia major patients may prevent or reverse ironoverload complications. *Blood Cells Mol Dis*. 2011;47(1):33–40.
54. Toumba M, Sergis A, Kanaris C, Skordis N. Endocrine complications in patients with Thalassaemia Major. *Pediatr Endocrinol Rev*. 2007;5(2):642–8.
55. Xia S, Zhang W, Huang L, Jiang H. Comparative efficacy and safety of deferoxamine, deferiprone and deferasirox on severe thalassemia: a meta-analysis of 16 randomized controlled trials. *PLoS One*. 2013;8(12).
56. Pakbaz Z, Fischer R, Treadwell M, Yamashita R, Fung E, Calvelli L, et al. A simple model to assess and improve adherence to iron chelation therapy with deferoxamine in patients with thalassemia. *Ann N Y Acad Sci*. 2005;1054:486–91.
57. Dasararaju R, Marques M. Adverse effects of transfusion. *Cancer Control*. 2015;22(1):16–25.
58. European Medicines Agency (EMA). Benefit-risk methodology project - Project description. 2009.

CHAPTER 8

59. Food and Drugs Administration (FDA). Structured Approach to Benefit-Risk Assessment in Drug Regulatory Decision-Making - Draft PDUFA V Implementation Plan. 2013.
60. Liberti B, McAuslane N, Walker S. Progress on the Development of a Benefit / Risk Framework for Evaluating Medicines. *Regul Focus*. 2010;15:32–7.
61. Coplan P, Noel R, Levitan B, Ferguson J, Mussen F. Development of a Framework for Enhancing the Transparency , Reproducibility and Communication of the Benefit – Risk Balance of Medicines. *Clin Pharmacol Ther*. 2011;89:312–5.
62. Walker S, McAuslane N, Liberti L, Salek S. Measuring benefit and balancing risk: strategies for the benefit-risk assessment of new medicines in a risk-averse environment. *Clin Pharmacol Ther*. 2009;85:241–6.
63. Contopoulos-Ioannidis D, Giotis N, Baliatsa D, Ioannidis J. Extended-interval aminoglycoside administration for children: a meta-analysis. *Pediatrics*. 2004;114(1):e111–8.

SECTION IV

Clinical Trial Simulations: accounting for exposure, disease progression and uncertainty in benefit-risk analysis

CHAPTER 9

Model-based evaluation of benefit-risk balance in children

F Bellanti, J Haddad, GC Del Vecchio, MC Putti, C Cosmi, A Ceci, A Maggio, J Horvath, M Danhof, and O Della Pasqua

Ready for submission

Summary

Aims: In this manuscript we apply a model-based approach to complement evidence generation and support an integrated evaluation of benefit-risk balance. Multicriteria decision analysis is used as a reference method for the benefit-risk analysis of chelation therapy for chronic iron overload in children. Thalassaemia was selected as a paradigm disease with the objective of assessing the impact of long term effects on the dose rationale for the paediatric population.

Methods: Clinical trial simulations and not-in-trial simulations were performed to characterise the time course of five clinical endpoints/markers deemed relevant for the evaluation of iron chelation therapy in paediatric patients affected by chronic iron overload. Simulations were based on hierarchical models previously developed using available clinical and literature data on deferoxamine. Summary statistics were used as input for multi-criteria decision analysis using the software D-Sight. For comparison purposes, deferoxamine, as a fixed dose of 45 mg/kg/day, was used as a reference scenario. A range of alternative dosing regimens and treatment follow-up periods up to 5 years were then evaluated, including fixed doses, weight-banded and ferritin-guided individualised regimens.

Results: The results of the MCDA show that fixed dosing regimens reach similar weighted scores in a typical phase III trial scenario. However the contribution of the different criteria varies considerably amongst the five endpoints. In addition, differences in the pharmacokinetics and pharmacodynamics of children below 20 kg and in patients with serum ferritin levels below 2500 µg/L suggest that these subgroups may benefit from alternative regimens. The differences in these groups appear to hold throughout the 5-year follow-up scenario, although the overall weighted scores decrease and the differences among treatment options are less evident.

Conclusions: In contrast to the evidence obtained during a phase III trial, the use of a model-based approach reveals that children below 20 kg and patients with ferritin levels below 2500 µg/L may achieve a similar BR score with higher and lower doses, respectively. Our analysis also shows the feasibility of integrating PKPD relationships into BR methodologies such as MCDA, allowing for a more clear, transparent and systematic assessment of the BRB of a medicinal product. Of relevance for paediatric diseases is the possibility to explore BRB beyond the duration of treatment in a clinical trial. Moreover, it illustrates how evidence synthesis can be complemented by simulated data, enabling the evaluation of options and scenarios which may not be available from empirical experimental protocols.

9.1 Introduction

Approval of new medicines for the paediatric population is based on the evidence regarding the efficacy and safety profile obtained throughout clinical development (1–4). However, a quantitative assessment of the benefit-risk balance (BRB) of a drug is usually not performed by sponsors or regulatory authorities at the moment of first marketing authorisation (5). Currently, quantitative assessment of the BRB remains a post-marketing endeavour, taking into account the emerging evidence from the therapeutic use of the drug in larger population and thereby mitigating some of the uncertainties associated with the limited data available at the time of launch. Interest towards the contribution of quantitative methodologies for BR assessment has increased considerably in the past years, with different stakeholders recognising the need for a more standardised framework, that includes higher transparency and consistency(6–14). Among the numerous approaches for quantitative BR analysis, it appears that the lack of transparency can be addressed by the development of multi-criteria decision analysis (MCDA) (6,14–21). Nonetheless, this and other methods rely on the assumption that a systematic review of empirical evidence arising from randomised clinical trials and observational studies data provides an accurate, unbiased picture of a drug's efficacy and safety.

This assumption though, may not be valid for a number of reasons. First, it should be noted that for many drugs the evidence required to support regulatory submission does not arise from the overall target population, as data is constrained by inclusion and exclusion criteria which may not be applicable during the therapeutic use of the medicinal product. In addition, little is done to disentangle the contribution of treatment on disease progression from external confounding factors on treatment response. Furthermore, the information collected in the context of pivotal clinical trials may not provide evidence that dose selection, dosing regimen, and treatment duration are truly optimal. Current approaches provide a solution to these issues only on the basis of data accumulation from larger clinical trials (before drugs approval is obtained) or from data obtained in post-marketing phases. In the past years, model-based drug development has proven to be an important resource in pharmaceutical research and may be an extremely helpful tool for projecting or hypothesising based on assumptions in anticipation of further data collection (22–25). Its value is particularly relevant in paediatric drug development where M&S can be used as a tool to characterise pharmacokinetic-pharmacodynamic relationships and support further understanding of the efficacy and safety profile of old and new drugs (22,24). In this manuscript, we propose a model-based approach to complement evidence generation for an integrated evaluation of BRB and provide an opportunity for a comprehensive evaluation before the first marketing approval. Chronic iron overload will be used as a paradigm disease with the objective of assessing the impact of long term effects on the dose rationale for the paediatric population.

Chronic iron overload is a consequence of chronic blood transfusions in patients affected by transfusion-dependent diseases such as beta-thalassaemia major (26–33). These patients experience a number of complications such as cardiac dysfunction, hypogonadism, hypothyroidism and diabetes mellitus due to tissue specific iron accumulation (27,28,30,32). In order to keep iron levels under control, these patients undergo therapy with iron chelators, which present a number of unfavourable effects, and along with disease-related complications affect the patients' quality of life (34). To provide an assessment of BRB as close as possible to clinical practice in this indication, we have selected deferoxamine as a reference compound. Deferoxamine is the currently considered as first line therapy for iron overload (34–36). However, we would like to stress that the context of the exercise is purely illustrative and is not intended to modify or provide recommendations about its benefit-risk profile.

Instead, our objective is to show how integration of modelling and simulation with quantitative methods such as MCDA can be used to complement evidence generation for diseases or conditions in which data arising from clinical development may be limited or insufficient to address clinical and regulatory questions at the time of marketing authorisation. We focus on the opportunities for incorporating pharmacokinetic-pharmacodynamic relationships into the evaluation of the dose rationale and reducing the uncertainty and empiricism in evidence synthesis during BR analyses.

9.2 Methods

Endpoints

All the data used in the analysis were simulated using pharmacokinetic, pharmacodynamic and disease models previously developed by our group. Five clinical endpoints were used for the evaluation of the BR framework for iron chelation therapy. A brief description of the selected efficacy and safety endpoints is provided below:

1. *Serum ferritin level* was selected as a measure of total body iron accumulation. Simulated data describing ferritin levels over time were included in the analysis as number of responders. A responder was defined as follows: a 20% reduction from baseline after 1 year of treatment for patients with baseline serum ferritin of 2500 µg/L or more; any decrease of serum ferritin levels or an increase, if that increase is less than 15% of the baseline as long as it does not result in levels above 2500 µg/L, for patients with baseline serum ferritin less than 2500 µg/L. Inclusion criteria at the start of treatment is described in the following paragraphs.
2. Hypothyroidism is a complication of the disease and its prevention was considered a benefit of the chelation therapy. Simulated data describing the *incidence of*

hypothyroidism was used as a measure of the progression of the disease. The reduction of its incidence is an overall favourable effect of drug therapy.

3. Diabetes mellitus is a complication of the disease and its prevention was considered a benefit of the chelation therapy. Simulated data describing the *incidence of diabetes* was used as a measure of the progression of the disease. The reduction of its incidence is an overall favourable effect of drug therapy.
4. Arthralgia and myalgia are a consequence of the chelation therapy by deferoxamine. This is a very common and dose-dependent AE of the iron chelator deferoxamine. It was simulated in terms of the *incidence of arthralgia/myalgia* in individual patients over the course of treatment.
5. Anaphylaxis is a rare dose-independent AE of the iron chelator deferoxamine. Simulated data reflected the *incidence of anaphylaxis* in individual patients. The occurrence of anaphylaxis would represent a drop-out from the study or switch to an alternative treatment, nonetheless, given the very low incidence patients' data were kept for the evaluation of the other endpoints.

The pharmacokinetic, pharmacodynamic and disease models were hierarchical models, with stochastic parameters describing within and between-subject variability. NONMEM v.7.2 and R software were used for simulation purposes as well as for graphical and statistical summaries. For the simulation of serum ferritin profiles a turnover model was previously built by our group, characterised by a disease model that accounts for the effect of the chronic transfusion regimen and by a drug model that accounts for the effect of iron chelators in reducing serum ferritin levels [Chapter 7 of this thesis]. For the simulation of the incidence of hypothyroidism and diabetes, two exponential hazard models were developed in which serum ferritin was included as a predictor of the instantaneous hazard [Chapter 8 of this thesis].

For the evaluation of drug-specific adverse events, a logistic model with nonlinear regression affected by changes in deferoxamine exposure was used to describe the incidence of arthralgia/myalgia in dose-dependent manner; whereas a truncated normal distribution was used in R to simulate anaphylaxis events in a dose-independent manner. 250 simulations were performed for each individual to account for inter- and intra-individual variability in the thalassemic population. An overview of the equations used to describe the response variable for each of the models is presented in the Table 1.

Table 1. Models used for the simulations

Model and equations	Description
<p style="text-align: center;"><i>Deferoxamine PK model</i></p> $\frac{dA(1)}{dt} = A(2) \times Q/V2 - A(1) \times Q/V1 - A(1) \times CL/V1$ $\frac{dA(2)}{dt} = A(1) \times Q/V1 - A(2) \times Q/V2$	<p>2 compartment PK model with zero-order absorption (8 hours subcutaneous infusion) and first-order elimination processes. Fixed allometric scaling (exponent of 0.75 on CL/F and 1 on V1/F and V2/F) is used to extrapolate exposure in adolescents and children</p>
<p style="text-align: center;"><i>Deferoxamine PKPD model</i></p> $\frac{dFERRITIN}{dt} = Kin + CRT - Kout \times FERRITIN \times (1 + DFO)$ $CRT = SCL \times e^{-SHP \times FERRITIN}$ $DFO = SLP \times SC_{SS}^{AV}$	<p>Kin = zero-order production rate Kout = first-order degradation rate CRT = disease component, additive production rate triggered by the transfusion regimen which was found to be non-linearly correlated to the disease status where SCL is a scaling factor and SHP is the shape factor of the correlation DFO = deferoxamine effect where SLP is the slope parameter of the concentration-effect relationship, and SC_{SS}^{AV} is the steady state concentrations</p>
<p style="text-align: center;"><i>Diabetes and Hypothyroidism hazard model</i></p> $S(t) = e^{-\int_0^t h(t)dt}$ $h(t) = h_0(t) * e^{\lambda frth + \lambda frtl}$	<p>The hazard is h(t), and the survival (S) is a function of the cumulative hazard within the time interval 0 to t. The effect of the disease is described by two components depending on whether serum ferritin levels are above (λfrth) or below (λfrtl) the threshold of 2500 µg/L</p>
<p style="text-align: center;"><i>Arthralgia/myalgia logistic model</i></p> $P = \frac{C_{SS}^\gamma}{(PC_{50}^\gamma + C_{SS}^\gamma)}$	<p>C_{SS} is the deferoxamine steady state concentration, PC₅₀ is the concentration corresponding to a 50% probability of experiencing the AE, and γ is the coefficient defining the shape of the relationship</p>

Phase III trial design

A phase III trial of the duration of 1 year was simulated in paediatric thalassaemic patients undergoing chelation therapy with deferoxamine at a fixed dose of 45 mg/kg/day for 5 days a week. A sample size of 150 patients was selected with about 30 patients aged 2 to 6 years, 70 aged 6 to 12 years and 50 aged 12 to 17 years. Patients' demographics were as follows (median and range): age 10 years (2-17), body weight 32 kg (12-62), 50% males and baseline ferritin levels 3000 µg/L (1000-8500). A graphical representation of the simulated serum ferritin profiles for the 1 year study is shown in Figure 1, whereas a summary of the remaining endpoints is presented in Table 2.

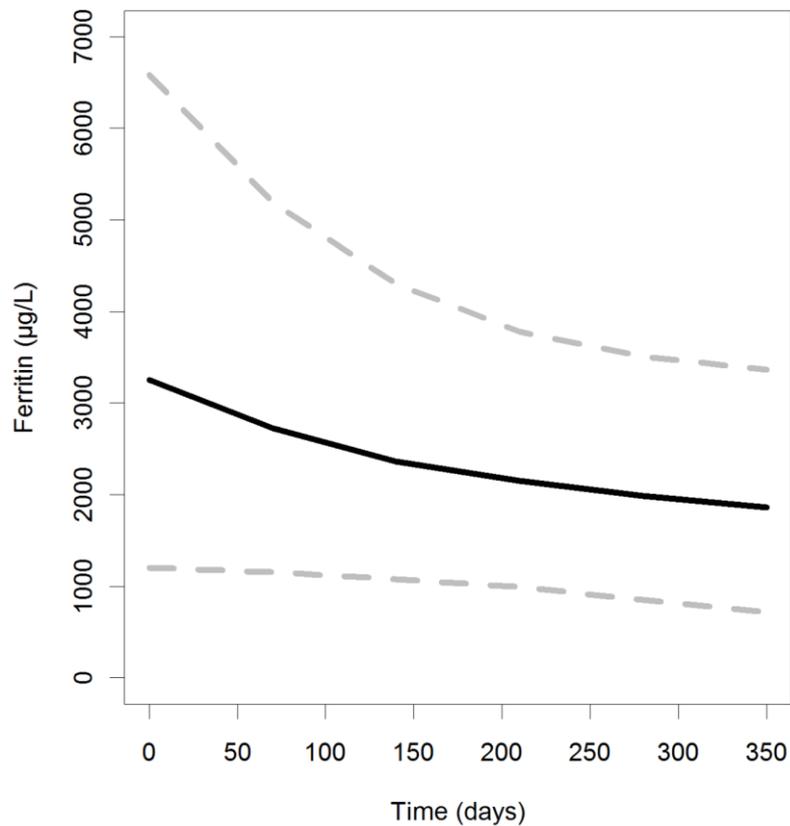


Figure 1. Simulated serum ferritin profiles over a period of 1 year for the Phase III trial in thalassaemic paediatric patients. The solid black line represents the median, whereas the dashed grey lines represent the 5th and 95th percentiles.

Table 2. Summary statistics of the simulated phase III trial

Endpoint	Units	Mean	LCI ¹	RCI ²	SD	5 P ³	95 P ⁴
Incidence of Hypothyroidism	%	3,26	3,07	3,45	1,54	0,67	6
Incidence of Diabetes	%	4,9	4,69	5,11	1,69	2,67	8
Incidence of Arthralgia/Myalgia	%	64,3	63,84	64,76	3,71	57,63	70
Incidence of Anaphylaxis	%	0,63	0,59	0,66	0,28	0,13	1,07

¹ Left confidence interval of the mean

² Right confidence interval of the mean

³ 5th percentile

⁴ 95th percentile

Complementary simulation scenarios

A number of alternative scenarios were simulated along with the phase III trial. A sample size of 150 patients (as commonly tested in phase III protocols for chronic iron overload in children) per treatment arm was selected also for the alternative scenarios. Even though a standard phase III trial in this patient population would last on average 1 year, we have simulated data for a period of 10 years to assess the changes in the long-term outcomes, with a number of 5 observations per year. Patients' demographics were similar to the one used for the phase III trial. In the end two scenarios were selected and used for the BR analysis, namely data simulated over a 1 year and a 5 year period. Summary statistics for the simulated data are presented in Tables 3 and 4 (see Appendix). Different dosing algorithms were tested and used as treatment options for the BR analysis; the different regimens are presented in Table 5. Along to the fixed dosing regimen of 45 mg/kg/day (5/7) used as a reference scenario (phase III trial), a range of different fixed doses were tested as well as individualised regimens based on body weights or serum ferritin differences.

Table 5. BR analysis scenarios

Input for standard MCDA analysis	Input for integrated PKPD and MCDA analysis	
Phase III data based on a fixed dose of 45 mg/kg/day 5/7	Scenario	Alternative options
	1: Fixed dosing regimens	30, 40, 50 and 60 mg/kg/day 5/7
	2: Weight banded dosing regimens	Kg < 20: 60 mg/kg/day 5/7
		20-40 kg: 50 mg/kg/day 5/7
		Kg > 40: 45 mg/kg/day 5/7
	3: Ferritin guided dosing regimens	Ferritin < 2500 µg/L: 40
		Ferritin 2500-5000 µg/L: 45
Ferritin > 5000 µg/L: 55		

Multi-criteria decision analysis

The MCDA analysis was performed with the software D-Sight (D-Sight Brussels, Belgium) which uses the PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluation) methods (37–40). The different stages of the analysis are summarised in Table 6 (17,18). Summary statistics of the simulated data discussed above were introduced in the MCDA software for the analysis. Mean and confidence intervals of the clinical endpoints for the different treatment arms and subgroups were used as input for the analysis (MCDA criteria).

Table 6. MCDA stages (adapted from Dogson et al.)

Stage	Description
1 - Establish the decision context	Establish aims of the MCDA, and consider the context of the appraisal
2 - Identify the options to be appraised	Define the options that will be evaluated in the appraisal
3 - Identify objectives and criteria	Identify criteria for assessing the consequences of each option and organise criteria into a value tree
4 - Scoring	Assess the expected performance of each option against the criteria; and assess the value associated with the consequences of each option for each criterion
5 - Weighting	Assign weights for each of the criterion to reflect their relative importance to the decision
6 - Derive an overall value	Calculate overall weighted score by combining weights and scores for each option
7 - Results	Examine the results and the contribution of individual criterion to the overall score
8 - Sensitivity analysis	Assess the influence of other preferences or weights on the overall ordering of the options

Expert input: value tree and weights elicitation

The analysis was conducted with a group of experts including: 1 former member of the PDCO (Paediatric Committee), 3 haematologists/paediatricians, 1 clinical trial expert, 1 statistician and 1 clinical pharmacologist.

Discussions with experts lead to the definition of the final value tree (a tree-like graph of the different criteria), as well as to the characterisation of the preference values for the criteria selected and the relative weights for the different criteria or weights elicitation (stages 4 and 5 of the MCDA analysis). The outcome of this process reflects the risk perception of the different stakeholders and has the objective of providing an adequate and unbiased risk assessment before the processing of the data is performed.

The final value tree is presented in Figure 2 and includes already the contribution of the relative weights assigned by the experts (weights elicitation), whereas Figure 3 shows an example of two utility functions defined for serum ferritin response (non-linear) and arthralgia/myalgia (linear) during the assessment of preference values.

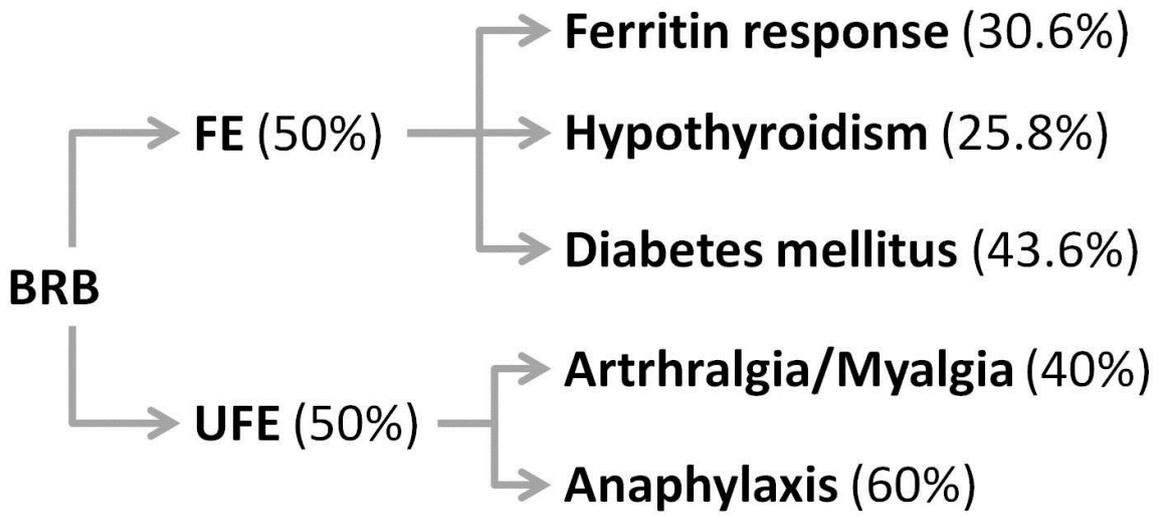


Figure 2. Final value tree and relative weights for the different criteria after discussion with experts. Favourable effects (FE) and unfavourable effects (UFE) were given the same importance whereas among the FE and UFE, diabetes and anaphylaxis were given the major importance respectively.

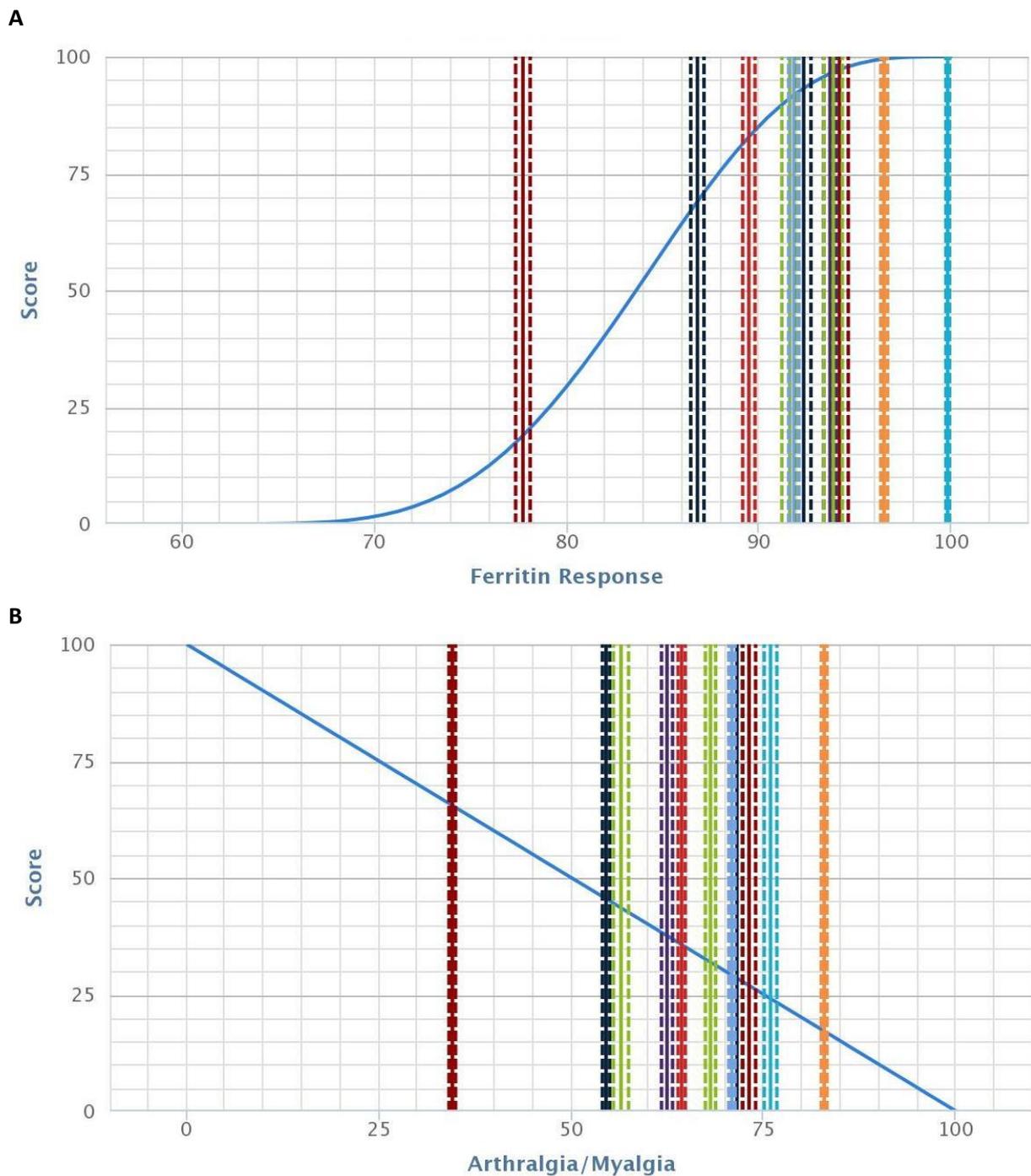


Figure 3. Assessed preference values for two criteria based on the discussion with experts. Panel A shows the non-linear utility function defined for ferritin response, whereas panel B shows the linear utility function defined for arthralgia/myalgia. On the y axis the score is presented in percentage (%).

With respect to the weights elicitation: as shown in Figure 2, the same importance was given to all favourable effects (FEs) against all unfavourable effects. Among the FEs, prevention of diabetes had a higher importance as compared to both ferritin response and prevention of

hypothyroidism; whereas the last two had equal importance as compared to each other. Finally, among the unfavourable effects (UFEs), greater relevance was given to anaphylaxis given the seriousness of the event. In addition, a brief summary of the discussion on the assessment of the preference values is provided below:

1. Ferritin response: a nonlinear increase was selected for this criterion reflecting an optimal response above 90% and a poor response below 80% as depicted in figure 3 (panel A).
2. Hypothyroidism: a linear decrease is expected to be sufficient to characterise the differences among the options under evaluation as hypothyroidism is considered relatively tolerable by the experts.
3. Diabetes mellitus: experts have defined an incidence above 5% as not acceptable. A non-linear utility function has been selected to characterise differences among the proposed options.
4. Arthralgia/Myalgia: a linear decrease (Figure 3, panel B) was considered sufficient to capture the differences among the options proposed as a high rate of the AE can still be tolerated according to the experts' opinion.
5. Anaphylaxis: a very steep non-linear decrease has been selected for anaphylaxis given the seriousness of the AE.

Calculation of the overall weighted score

With the information on preference values and relative weights, the final step was to calculate the overall weighted score for each option (stage 6 of MCDA). The outcome of this calculation is simply the weighted average of its scores on the different criteria. The final score is generated using the following equation:

$$S_i = w_1s_{i1} + w_2s_{i2} + \dots + w_ns_{in} \quad \text{Equation 1}$$

where the overall weighted score (S) for an option i will be given by the sum of all the individual scores (s) of each criterion multiplied by the assigned weight (w).

Assumptions

We assume that the incidence of these effects is not random, in contrast to current approaches that regard the various endpoints as independent of each other. We captured mechanistic correlations across the various endpoints as described in the equations of table 1, except for anaphylaxis which is a dose-independent AE. For the evaluation of unfavourable effects we have selected frequency as the only dimension of interest for this analysis, without taking into account severity or duration, i.e. assuming a grade 2-3 for all

AEs. We recognise that in clinical practice, severity and duration have an essential role in the evaluation of the BR balance and therefore should be accounted for. Furthermore, when a fixed dosing regimen was evaluated during the 1 year trial we maintained a fixed regimen also during the follow-up years and in the same manner, independently of patients' response to therapy, no switch therapy was considered. On top of that, treatment compliance was assumed to be optimal in this exercise and subsequently the observed differences are essentially due to variability in pharmacokinetics. We acknowledge the importance of these factors, nonetheless, we chose to reduce the complexity to better illustrate the advantage of the approach without influencing its validity.

9.3 Results

The results of the multi-criteria decision analysis are presented in Figures 4 and 5, for the 1 year clinical trial and 5 year treatment follow-up, respectively. Figure 4 shows that the fixed dosing regimens have a similar weighted score; except for the 30 mg/kg regimen (score of 29.28) where the lowest score is achieved. Even though the overall score is similar the contribution of the different criteria is differs considerably amongst the five endpoints, with, as expected, ferritin response that tends to increase at increasing doses counteracted by the contribution of AEs that tends to increase as the dose decreases.

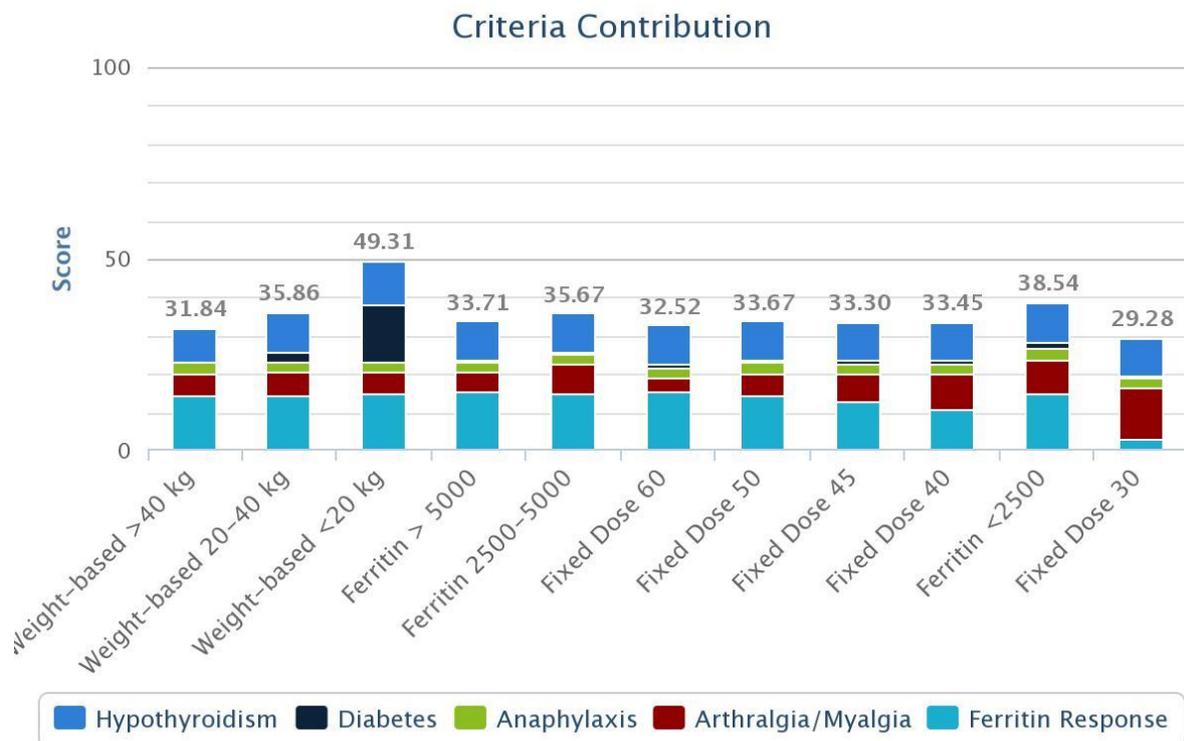


Figure 4. Criteria contribution for the 1 year scenario. The overall weighted score is presented for the different options (the higher the score the better the overall performance of the option appraised). Individual criteria contribution are displayed for each option: light blue, dark red, green, dark blue and blue represent respectively ferritin response, arthralgia/myalgia, anaphylaxis, diabetes and hypothyroidism.

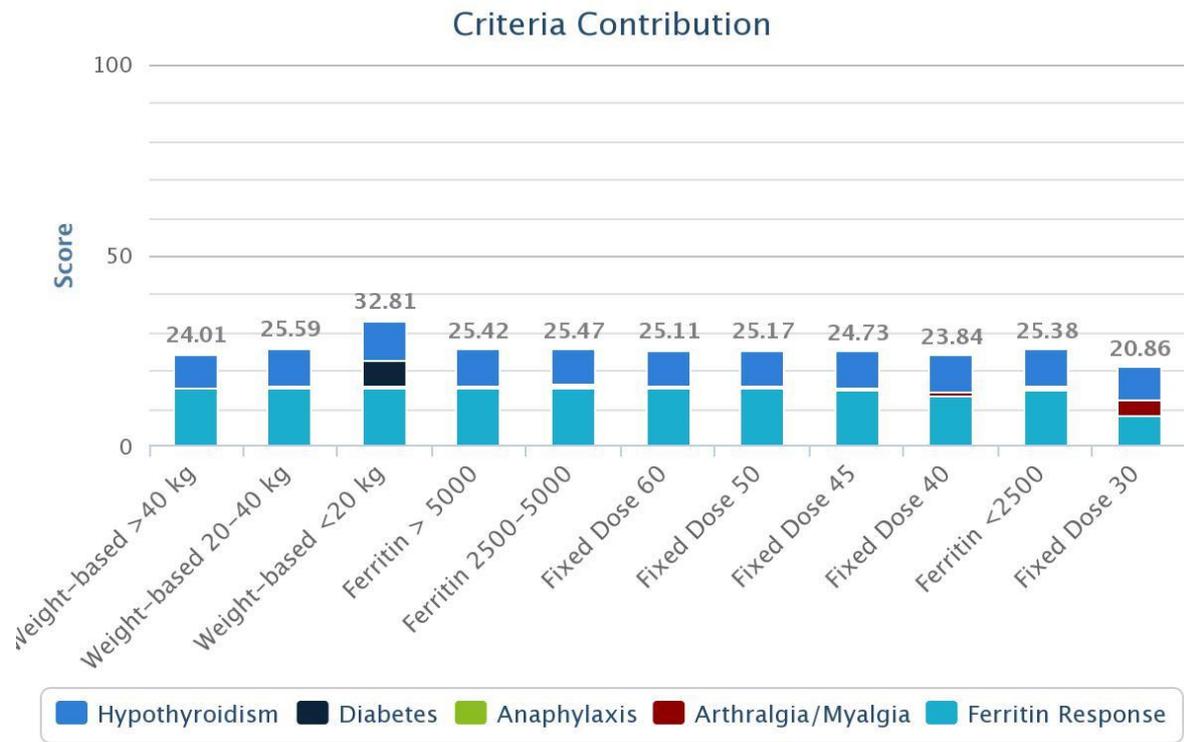


Figure 5. Criteria contribution for the 5 year scenario. The overall weighted score is presented for the different options (the higher the score the better the overall performance of the option appraised). Individual criteria contribution are displayed for each option: light blue, dark red, green, dark blue and blue represent respectively ferritin response, arthralgia/myalgia, anaphylaxis, diabetes and hypothyroidism.

The results of the MCDA show that fixed dosing regimens reach similar weighted scores in a typical phase III trial scenario. However the contribution of the different criteria varies considerably amongst the five endpoints. In addition, differences in the pharmacokinetics and pharmacodynamics of children below 20 kg and in patients with serum ferritin levels below 2500 µg/L suggest that these subgroups may benefit from alternative regimens. The differences in these groups appear to hold throughout the 5-year follow-up scenario, although the overall weighted scores decrease and the differences among treatment options are less evident. From the 5-year treatment follow up is also clear that the acute effects become clinically less relevant; in addition, the differences among the individual contributions of each criterion tend to disappear. For example, in a five year period different

doses lead to a similar response in serum ferritin. Yet, such changes are achieved at very different rates.

9.4 Discussion and conclusion

MCDA results

Before any quantitative BR evaluation is performed, the integration of multiple models is essential and allows to 1) complementing the existing data to support the decision to be taken and possibly determining whether personalised medicine would be of any benefit for the patient population; 2) optimising the input data for the BR analysis; and 3) quantifying the relevant correlations among different endpoints that are currently still evaluated in an independent manner.

Assuming that the scenarios presented here are part of a real clinical case, the therapeutic conclusion derived from this analysis may be the following: children below 20 kg may benefit from a higher dose (60 mg/kg/day) at least in an early phase of the disease, and patients with controlled serum ferritin levels below 2500 µg/L may achieve a similar BR score with a lower dose (40 mg/kg/day), as compared to the evidence arising from the phase III trial data (fixed 45 mg/kg/day). A model-based approach allows one to understand the implications of doses that have not been formally tested and the impact they have on benefit and risk. The approach also enables one to take into account clinical and feasibility elements that were not considered in the clinical protocols. In addition, the possibility to explore beyond the standard duration of a phase III trial allows understanding how long-term outcomes may affect the BR scores and anticipate whether any changes can be expected in the BR balance of the drug.

Limitations

It is important to emphasise that it was not our intent to modify in any way the current BR balance of deferoxamine; our goal was to demonstrate how model-based MCDA can be used to personalise drug therapy by incorporating various alternatives and virtual sub-populations in the analysis. The complexity of chronic iron overload is much higher than the one depicted in this manuscript in many ways: e.g., other disease-related complications, such as cardiac complications, have a higher relevance in the evaluation of iron accumulation; drop-out rates that occur in a real clinical setting have not been considered during this analysis; and last but not least the role that treatment compliance (especially for deferoxamine) has on the clinical evaluation of iron overload is extremely important. Having acknowledged that, an exercise with less complexity provides a better framework for illustrating how modelling and simulation can be used to overcome some of the issues highlighted in the manuscript.

Even though in the recent years PKPD modelling has been proposed in conjunction with clinical utility approaches (41,42), in this manuscript we integrate for the first time PKPD modelling with multi-criteria decision analysis (MCDA) to overcome the issues discussed in the introduction.

Furthermore, we learned from this exercise that given the complexity that usually characterises the BR evaluation of drugs, a quantitative and integrated approach is essential to reduce the uncertainty of the analysis and to increase the understanding of the BRB. This is particularly true in the paediatric context where not only the BRB is not constant over time (in particular in chronic diseases, as in the example discussed here), but also the lack of available data does not allow performing an appraisal that is representative of real life population (22–24,43,44). Complementing evidence generation (i.e., real data) with virtual scenarios and alternative treatment and protocol options (clinical and feasibility elements such as study design, inclusion and exclusion criteria, etc.) using clinical trial simulations and/or not-in-trial simulations provides an opportunity to accomplish two major goals: achieving a better and more comprehensive understanding of the BRB possibly before a drug reaches the market and evaluating the BRB in sub-groups providing the basis for the assessment of personalised therapy. This is an element often overlooked in that understanding of BRB is also relevant for children, their parents and others interested in patients engagement.

Conclusion

In conclusion, we have successfully complemented evidence generation using PKPD modelling to the use of MCDA for BR assessment in a paediatric disease. We strongly believe that such an approach is essential for a more structured evaluation of the BR balance of any intervention, especially if mechanism-based modelling and pharmacokinetic-pharmacodynamic relationships are used to support such scenarios. Of relevance for paediatric diseases is the possibility to explore BRB beyond the duration of treatment in a clinical trial.

References

1. Downing NS, Aminawung J a, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA*. 2014;311(4):368–77.
2. Food and Drug Administration (FDA). Guidance for Industry Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products. 1998;(May):1–23.
3. Food and Drug Administration (FDA). Guidance for Industry: Exposure-Response Relationships - Study Design, Data Analysis and Regulatory Applications. 2003;(April):1–25.
4. Katz R. FDA: evidentiary standards for drug development and approval. *NeuroRx*. 2004;1(July):307–16.
5. European Medicines Agency (EMA), Committee for medicinal products for human use (CHMP). Guideline on safety and efficacy follow-up - risk management of advanced therapy medicinal products. 2008.
6. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. *Value Health* 2010;13:657–66.
7. Holden WL. Benefit-risk analysis : a brief review and proposed quantitative approaches. *Drug Saf* 2003;26:853–62.
8. Cohen J, Neumann P. What's more dangerous, your aspirin or your car? Thinking rationally about drug risks (and benefits). *Heal Aff*. 2007;26:636–46.
9. Bjornson D. Interpretation of drug risk and benefit: individual and population perspectives. *Ann Pharmacother*. 2004;38:694–9.
10. EMA. Benefit-risk methodology project - Project description. 2009.
11. FDA. Structured Approach to Benefit-Risk Assessment in Drug Regulatory Decision-Making - Draft PDUFA V Implementation Plan. 2013.
12. Liberti B, McAuslane N, Walker S. Progress on the Development of a Benefit / Risk Framework for Evaluating Medicines. *Regul Focus*. 2010;15:32–7.
13. Coplan P, Noel R, Levitan B, Ferguson J, Mussen F. Development of a Framework for Enhancing the Transparency , Reproducibility and Communication of the Benefit – Risk Balance of Medicines. *Clin Pharmacol Ther*. 2011;89:312–5.
14. Walker S, McAuslane N, Liberti L, Salek S. Measuring benefit and balancing risk: strategies for the benefit-risk assessment of new medicines in a risk-averse environment. *Clin Pharmacol Ther*. 2009;85:241–6.

CHAPTER 9

15. EMA. Benefit-risk methodology project - work package 2 report: Applicability of current tools and processes for regulatory benefit-risk assessment. 2011.
16. EMA. Benefit-risk methodology project Work package 4 report : Benefit-risk tools and processes. 2012.
17. Keeney RL, Raiffa H. Decisions with Multiple Objectives. Cambridge: Cambridge University Press; 1993. 1-569 p.
18. Dogson. Multi-criteria analysis: a manual. Communities and Local Government Publications; 2009.
19. Mussen F, Salek S, Walker S. Benefit-Risk Appraisal of Medicines. A systematic approach to decision-making. 2009.
20. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines – part 1 : The development of a new model using multi-criteria decision analysis. *Pharmacoepidemiol Drug Saf.* 2007;(June):2–15.
21. Walker S, Cone M. The Development of a Model for Benefit-Risk Assessment of Medicines Based on Multi-Criteria Decision Analysis (Summary Report). 2004.
22. Manolis E, Osman T, Herold R, Koenig F, Tomasi P, Vamvakas S, et al. Role of modeling and simulation in pediatric investigation plans. *Paediatr Anaesth.* 2011;21:214–21.
23. Manolis E, Pons G. Proposals for model-based paediatric medicinal development within the current European Union regulatory framework. *Br J Clin Pharmacol.* 2009;68:493–501.
24. Bellanti F, Della Pasqua O. Modelling and simulation as research tools in paediatric drug development. *Eur J Clin Pharmacol.* 2011;67(Suppl 1):75–86.
25. Danhof M, de Jongh J, De Lange E, Della Pasqua O, Ploeger B, Voskuyl R. Mechanism-based pharmacokinetic-pharmacodynamic modeling: biophase distribution, receptor theory, and dynamical systems analysis. *Annu Rev Pharmacol Toxicol.* 2007;47:357–400.
26. Andrews N. Disorders of Iron Metabolism. *N Engl J Med.* 1999;341(26):1986–95.
27. Borgna-Pignatti C, Cappellini MD, De Stefano P, Del Vecchio GC, Forni GL, Gamberini MR, et al. Survival and complications in thalassemia. *Ann N Y Acad Sci.* 2005 Jan;1054:40–7.
28. Borgna-Pignatti C, Rugolotto S, De Stefano P, Zhao H, Cappellini MD, Del Vecchio G, et al. Survival and complications in patients with thalassemia major treated with transfusion and deferoxamine. *Haematologica.* 2004;89(10):1187–93.
29. Borgna-Pignatti C, Cappellini MD, Stefano P De, Carlo G, Vecchio D, Forni GL, et al. Cardiac morbidity and mortality in deferoxamine- or deferiprone-treated patients with thalassemia major. 2009;107(9):3733–7.

30. Cunningham MJ, Macklin E a, Neufeld EJ, Cohen AR. Complications of beta-thalassemia major in North America. *Blood*. 2004 Jul 1;104(1):34–9.
31. Galanello R, Origa R. Beta-thalassemia. *Orphanet J Rare Dis*. 2010 Jan;5:11.
32. Rund D, Rachmilewitz E. Beta-Thalassemia. *N Engl J Med*. 2005;353:1135–46.
33. TIF. Guidelines for the clinical management of thalassaemia. 2008.
34. Novartis Pharmaceuticals UK. Deferoxamine summary of product characteristics. Available from: <http://www.medicines.org.uk/emc/medicine/2666>
35. Fisher S, Brunskill S, Doree C, Gooding S, Chowdhury O, Roberts D. Desferrioxamine mesylate for managing transfusional iron overload in people with transfusion-dependent thalassaemia (Review). *cochrane Libr*. 2013;(8).
36. Porter J. Deferoxamine pharmacokinetics. *Semin Hematol*. 2001 Jan;38(1 Suppl 1):63–8.
37. Brans J, De Smet Y. Chapter 1 PROMETHEE METHODS. In: Figueira J, Greco S, Ehrgott M, editors. *Multiple criteria decision analysis: state of the art surveys*. p. 1–44.
38. Aenishaenslin C, Hongoh V, Cissé HD, Hoen AG, Samoura K, Michel P, et al. Multi-criteria decision analysis as an innovative approach to managing zoonoses: results from a study on Lyme disease in Canada. *BMC Public Health*. 2013 Jan;13:897.
39. Mutikanga HE, Sharma SK, Vairavamoorthy K. Multi-criteria Decision Analysis: A Strategic Planning Tool for Water Loss Management. *Water Resour Manag*. 2011 Aug 5;25(14):3947–69.
40. Cassidy D, Thomson G, Barnes J. Establishing Priorities through use of Multi-criteria Decision Analysis for a Commodity Based Trade approach to Beef Exports from the East Caprivi Region of Namibia. 2013.
41. Ouellet D, Werth J, Parekh N, Feltner D, McCarthy B, Lalonde RL. The use of a clinical utility index to compare insomnia compounds: a quantitative basis for benefit-risk assessment. *Clin Pharmacol Ther*. 2009 Mar;85(3):277–82.
42. Ouellet D. Benefit-risk assessment: the use of clinical utility index. *Expert Opin Drug Saf*. 2010 Mar;9(2):289–300.
43. EMA. Paediatric Workshops - http://www.emea.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000416.jsp&mid=WC0b01ac0580118a19.
44. ICH Topic E11. http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002926.pdf 2001.

Appendix

Table 3. Summary statistics of the simulated data for the 1 year scenario

Option	Criteria	Units	Mean	LCI ¹	RCI ²	SD	5 P ³	95 P ⁴
Phase III Fixed dose 45	Ferritin response	%	89,44	89,12	89,76	2,56	85,33	93,33
	Incidence of Hypothyroidism	%	3,26	3,07	3,45	1,54	0,67	6
	Incidence of Diabetes	%	4,9	4,69	5,11	1,69	2,67	8
	Incidence of Arthralgia/Myalgia	%	64,3	63,84	64,76	3,71	57,63	70
	Incidence of Anaphylaxis	%	0,63	0,59	0,66	0,28	0,13	1,07
Fixed dose 30	Ferritin response	%	77,7	77,32	78,08	3,05	72,67	82,67
	Incidence of Hypothyroidism	%	3,33	3,13	3,52	1,56	0,67	6
	Incidence of Diabetes	%	5,02	4,81	5,22	1,66	2,67	7,33
	Incidence of Arthralgia/Myalgia	%	34,5	34,08	34,92	3,36	28,67	39,7
	Incidence of Anaphylaxis	%	0,67	0,62	0,69	0,29	0,19	1,14
Fixed dose 40	Ferritin response	%	86,77	86,42	87,11	2,79	82,67	91,03
	Incidence of Hypothyroidism	%	3,18	2,99	3,37	1,5	0,67	6
	Incidence of Diabetes	%	4,75	4,54	4,96	1,66	2	7,7
	Incidence of Arthralgia/Myalgia	%	54,47	54	54,95	3,82	48	60,67
	Incidence of Anaphylaxis	%	0,67	0,59	0,65	0,27	0,13	1,06
Fixed dose 50	Ferritin response	%	91,78	91,51	92,05	2,18	88	95,33
	Incidence of Hypothyroidism	%	3,13	2,94	3,31	1,5	0,67	5,33
	Incidence of Diabetes	%	4,77	4,55	4,98	1,71	2	7,33
	Incidence of Arthralgia/Myalgia	%	70,83	70,43	71,24	3,28	64,67	76
	Incidence of Anaphylaxis	%	0,66	0,62	0,69	0,28	0,27	1,07
Fixed dose 60	Ferritin response	%	96,47	96,3	96,65	1,44	94	98,67
	Incidence of Hypothyroidism	%	3,19	3,01	3,38	1,52	0,67	6

MODEL-BASED EVALUATION OF BENEFIT-RISK BALANCE IN CHILDREN

	Incidence of Diabetes	%	4,71	4,51	4,92	1,69	2	7,33
	Incidence of Arthralgia/Myalgia	%	82,82	82,44	83,19	3,02	78	87,33
	Incidence of Anaphylaxis	%	0,65	0,62	0,68	0,28	0,26	1,14
Weight < 20 kg	Ferritin response	%	94,15	93,7	94,61	3,65	88,24	100
	Incidence of Hypothyroidism	%	1,74	1,47	2,01	2,18	0	5,88
	Incidence of Diabetes	%	2,42	2,12	2,73	2,47	0	5,88
	Incidence of Arthralgia/Myalgia	%	73,06	72,21	73,91	6,88	61,76	85,29
	Incidence of Anaphylaxis	%	0,66	0,62	0,69	0,29	0,27	1,20
Weight 20-40 kg	Ferritin response	%	91,57	91,16	91,97	3,27	85,25	96,72
	Incidence of Hypothyroidism	%	2,72	2,46	2,98	2,11	0	6,56
	Incidence of Diabetes	%	4,23	3,9	4,56	2,68	0	8,2
	Incidence of Arthralgia/Myalgia	%	68,05	67,38	68,72	5,41	60,66	77,05
	Incidence of Anaphylaxis	%	0,67	0,63	0,71	0,30	0,27	1,20
Weight > 40 kg	Ferritin response	%	92,29	91,91	92,67	3,05	87,27	96,36
	Incidence of Hypothyroidism	%	4,47	4,1	4,83	2,91	0	9,09
	Incidence of Diabetes	%	6,47	6,07	6,86	3,19	1,82	10,91
	Incidence of Arthralgia/Myalgia	%	71,49	70,77	72,22	5,85	61,82	81
	Incidence of Anaphylaxis	%	0,62	0,59	0,66	0,28	0,13	1,07
Ferritin < 2500	Ferritin response	%	93,83	93,35	94,3	3,83	86,49	100
	Incidence of Hypothyroidism	%	3,06	2,67	3,45	3,13	0	8,11
	Incidence of Diabetes	%	4,29	3,9	4,69	3,19	0	10,81
	Incidence of Arthralgia/Myalgia	%	56,42	55,44	57,4	7,93	43,24	67,57
	Incidence of Anaphylaxis	%	0,64	0,61	0,68	0,28	0,27	1,20
Ferritin 2500-	Ferritin response	%	93,66	93,33	93,99	2,65	88,75	97,37
	Incidence of Hypothyroidism	%	3,31	3,06	3,55	2	0	6,58

5000	Incidence of Diabetes	%	4,94	4,61	5,26	2,64	1,32	9,21
	Incidence of Arthralgia/Myalgia	%	62,4	61,68	63,12	5,82	53,22	71,05
	Incidence of Anaphylaxis	%	0,65	0,61	0,68	0,28	0,26	1,60
Ferritin > 5000	Ferritin response	%	99,77	99,68	99,87	0,75	97,3	100
	Incidence of Hypothyroidism	%	3,03	2,67	3,38	2,86	0	8,11
	Incidence of Diabetes	%	4,99	4,55	5,44	3,59	0	10,81
	Incidence of Arthralgia/Myalgia	%	75,85	75,03	76,67	6,6	64,86	86,49
	Incidence of Anaphylaxis	%	0,65	0,62	0,68	0,27	0,27	1,07

¹ Left confidence interval of the mean

² Right confidence interval of the mean

³ 5th percentile

⁴ 95th percentile

Table 4. Summary statistics of the simulated data for the 5 years scenario

Option	Criteria	Units	Mean	LCI ¹	RCI ²	SD	5 P ³	95 P ⁴
Phase III Fixed dose 45	Ferritin response	%	93,32	93,06	93,58	2,08	89,63	96,67
	Incidence of Hypothyroidism	%	4	3,78	4,22	1,74	1,33	7,33
	Incidence of Diabetes	%	5,82	5,59	6,05	1,84	3,33	9,33
	Incidence of Arthralgia/Myalgia	%	97,13	96,97	97,29	1,32	94,67	99,33
	Incidence of Anaphylaxis	%	0,66	0,65	0,68	0,13	0,45	0,87
Fixed dose 30	Ferritin response	%	83,97	83,65	84,3	2,61	79,33	88,37
	Incidence of Hypothyroidism	%	4,57	4,34	4,81	1,87	1,63	8
	Incidence of Diabetes	%	6,61	6,37	6,84	1,93	3,33	10
	Incidence of Arthralgia/Myalgia	%	80,6	80,22	80,97	3,04	75,33	85,33
	Incidence of Anaphylaxis	%	0,64	0,63	0,66	0,13	0,45	0,85
Fixed dose 40	Ferritin response	%	90,24	89,97	90,5	2,13	86,67	93,33
	Incidence of Hypothyroidism	%	4,04	3,83	4,25	1,72	1,33	7,33

MODEL-BASED EVALUATION OF BENEFIT-RISK BALANCE IN CHILDREN

	Incidence of Diabetes	%	5,94	5,71	6,18	1,89	3,33	9,33
	Incidence of Arthralgia/Myalgia	%	94,11	93,89	94,33	1,77	91,33	96,67
	Incidence of Anaphylaxis	%	0,65	0,63	0,66	0,12	0,45	0,86
Fixed dose 50	Ferritin response	%	95,44	95,24	95,64	1,6	92,67	98
	Incidence of Hypothyroidism	%	3,83	3,61	4,04	1,73	1,33	6,67
	Incidence of Diabetes	%	5,61	5,38	5,84	1,83	2,67	8,67
	Incidence of Arthralgia/Myalgia	%	98,24	98,12	98,36	0,97	96,67	100
	Incidence of Anaphylaxis	%	0,67	0,65	0,68	0,14	0,44	0,91
Fixed dose 60	Ferritin response	%	98,2	98,07	98,33	1,04	96,67	100
	Incidence of Hypothyroidism	%	3,82	3,61	4,04	1,72	1,33	6,67
	Incidence of Diabetes	%	5,55	5,33	5,78	1,82	2,67	8,67
	Incidence of Arthralgia/Myalgia	%	99,59	99,53	99,66	0,53	98,67	100
	Incidence of Anaphylaxis	%	0,66	0,64	0,67	0,14	0,45	0,93
Weight < 20 kg	Ferritin response	%	96,05	95,65	96,45	3,24	91,18	100
	Incidence of Hypothyroidism	%	2,62	2,29	2,96	2,73	0	7,5
	Incidence of Diabetes	%	3,41	3,03	3,8	3,09	0	8,82
	Incidence of Arthralgia/Myalgia	%	98,38	98,13	98,63	2,02	94,12	100
	Incidence of Anaphylaxis	%	0,66	0,64	0,68	0,13	0,45	0,90
Weight 20-40 kg	Ferritin response	%	94,67	94,33	95,01	2,77	90,16	98,36
	Incidence of Hypothyroidism	%	3,39	3,1	3,68	2,35	0	8,2
	Incidence of Diabetes	%	5,28	4,92	5,63	2,87	1,64	9,84
	Incidence of Arthralgia/Myalgia	%	98,09	97,87	98,3	1,7	95,08	100
	Incidence of Anaphylaxis	%	0,64	0,63	0,66	0,13	0,43	0,88
Weight > 40 kg	Ferritin response	%	95,24	94,89	95,58	2,79	90,91	100
	Incidence of Hypothyroidism	%	5,03	4,65	5,42	3,11	0	10,91

	Incidence of Diabetes	%	7,27	6,85	7,69	3,41	1,82	12,73
	Incidence of Arthralgia/Myalgia	%	98,36	98,15	98,58	1,73	94,55	100
	Incidence of Anaphylaxis	%	0,65	0,64	0,67	0,13	0,45	0,88
Ferritin < 2500	Ferritin response	%	92,64	92,12	93,15	4,14	86,49	98,78
	Incidence of Hypothyroidism	%	3,56	3,16	3,96	3,23	0	8,11
	Incidence of Diabetes	%	4,96	4,53	5,39	3,46	0	10,81
	Incidence of Arthralgia/Myalgia	%	96,86	96,49	97,24	2,99	91,89	100
	Incidence of Anaphylaxis	%	0,65	0,63	0,66	0,14	0,45	0,88
Ferritin 2500-5000	Ferritin response	%	97,99	97,8	98,19	1,59	94,74	100
	Incidence of Hypothyroidism	%	3,97	3,68	4,26	2,34	0	7,89
	Incidence of Diabetes	%	5,66	5,32	6	2,73	1,32	10,53
	Incidence of Arthralgia/Myalgia	%	96,96	96,72	97,19	1,89	93,42	100
	Incidence of Anaphylaxis	%	0,65	0,64	0,67	0,14	0,43	0,88
Ferritin > 5000	Ferritin response	%	99,95	99,9	99,99	0,38	100	100
	Incidence of Hypothyroidism	%	3,75	3,36	4,14	3,12	0	8,11
	Incidence of Diabetes	%	5,96	5,47	6,45	3,95	0	13,51
	Incidence of Arthralgia/Myalgia	%	97,99	97,73	98,25	2,08	94,59	100
	Incidence of Anaphylaxis	%	0,66	0,64	0,67	0,13	0,43	0,85

¹ Left confidence interval of the mean

² Right confidence interval of the mean

³ 5th percentile

⁴ 95th percentile

SECTION V

Conclusion and Perspectives

CHAPTER 10

Model-informed assessment of the benefit-risk profile of medicines for children

Summary, conclusions and perspectives

Growing awareness about the relevance of formal evaluation of the efficacy and safety in children has resulted into important changes in the legislation defining the requirements for the approval of medicines for children (1–4). In parallel to these developments, methodological advancements have taken place in terms of the level and type of evidence required to establish the so-called benefit-risk profile of an intervention (5–8). Whilst a considerable number of approaches have been evaluated over the last decade, their utilisation has often been limited to post-approval data. Most importantly, they summarise a *fait accompli*, i.e., the evidence is gathered after the facts.

Whilst risk management and mitigation measures are intrinsic components of a risk management plans (5–8), current approaches do not provide a quantitative framework for regulators, clinical scientists and drug developers on how to integrate knowledge about drug- and disease-specific properties, thereby enabling the prediction of treatment response across a range of possible scenarios before evidence is generated. The availability of such a framework would not only permit optimisation of risk management plans, it would also represent a more robust basis for addressing clinical and scientific questions during drug development and at the time of approval.

Throughout this thesis we have focused on the advantages of introducing quantitative clinical pharmacology concepts, and more specifically modelling and simulation, as an ancillary tool for evidence generation and evidence synthesis. We have illustrated how model-based predictions can be used in conjunction with established benefit-risk methodologies to support the decision-making process underpinning the approval of paediatric medicines. The examples used in previous chapters also offer insight into the deficiencies associated with data generation and unravel opportunities for the optimisation of clinical protocols in children.

Two main features need to be highlighted, which differentiate the work proposed here from previous research in paediatric clinical pharmacology. In contrast to previous work in which population pharmacokinetic-pharmacodynamic models have been developed to describe a

single endpoint, it is the first time that multiple drug-disease models are implemented in parallel, taking into account eventual correlations between measures of efficacy and safety. This represents an important advancement in the way one assesses treatment response i.e., not as a primary endpoint in a clinical protocol, but rather as a means to characterise disease- from drug-specific properties, thereby providing ***a parametric representation of the efficacy and safety profile of an intervention***. A second feature of our work is the application of clinical trial and not-in-trial simulations as complement to data obtained from clinical trials. Here simulated data (i.e., imputations) from virtual scenarios were intertwined with real data and used as input for the multi-criteria decision analysis. An immediate advantage of the approach is the possibility of ***exploring in a quantitative manner the benefit-risk profile of a medicinal product in situations which have not been tested prior to its approval***. This aspect is particularly relevant for the evaluation of medicines for children, for whom limited evidence can be generated and physiological processes associated with maturation and growth may affect the benefit-risk balance.

The aforementioned features were embedded across the different chapters, where chelation therapy associated with iron overload is used to illustrate the implementation of the proposed framework. Here we present an overview of the results and conclusions from these investigations, emphasising the contribution of modelling and simulation as a tool for more effective data generation, evidence synthesis and decision making regarding the evaluation of paediatric medicines.

Our work is based on the premise that when a drug is granted its first marketing authorisation the decision is based only on the evidence generated throughout the drug development phases in the target paediatric population (1,3,2,4). However, at this stage no quantitative evaluation is performed of the benefit-risk balance (BRB); usually a full BR appraisal takes place during the post-marketing phase, when additional evidence arises from clinical practice as well as from additional randomised controlled trials. Clearly, this situation is not ideal, as it imposes a reactive rather pro-active attitude towards benefit-risk. Despite the acknowledgement by different stakeholders about the need for a more consistent, transparent framework to support (decision-making for) the approval of new medicines (9–12), inferential methods by modelling and simulation have been ignored or have limited role as a statistical analysis tool. Thus far little has been done to enable the use of inferential methods by modelling and simulation as an integrative tool for evidence synthesis and benefit-risk assessment.

In **chapter 1**, we review the available literature on benefit-risk evaluation to identify suitable methods for the development of the proposed framework. In spite of the vast number of

methodologies (both qualitative and quantitative) are available in the public domain, the majority of them are not appropriate for a more general application (13–15).

Among other things, we highlight the relevance of quantitative methods as enablers or keys to the answer to clinical, regulatory and scientific questions regarding the benefits and risks of an intervention. Growing consensus suggests that a combined approach involving qualitative and quantitative methods is required to ensure meaningful evaluation and interpretation of benefit and risk data. Here we identify MCDA as the method of choice for further integration with mechanism-based modelling and simulation. Despite its limitations in the way uncertainty is handled, MCDA offers the opportunity to evaluate a multidimensional aspects drug and disease which arise in drug development and in the clinical practice. In revisiting the drug approval process and the requirements for paediatric drug development, it becomes evident that the use of drug-disease modelling and simulation represents a formal extension of the clinical pharmacology concepts into the realm of evidence synthesis and evaluation of novel therapeutic agents. This advancement can be compared to the introduction of receptor pharmacology in drug discovery, which replaced empirical evidence from experimental protocols (16–18). Then receptors were just a concept, not a substrate, whose properties could be used to understand drug properties and optimise the development of novel molecules. Similarly, today response scenarios in virtual patients are still seen as concepts, rather than as substrates that can be used for further characterisation of the benefit-risk profile.

Having identified a suitable methodology enabled us to formalise the scope and intent of the investigations described in the subsequent chapters of this thesis. In fact, in **chapter 2** we introduce details on the implementation of a framework in which MCDA is applied in an integrated manner with modelling and simulation. The primary intent of the framework is to have a tool for more effective data generation, evidence synthesis and better decision making. Focus is given to the opportunities for optimising data generation in children and most importantly to the possibility of integrating knowledge by mechanism-based parameterisations, which enable us to discriminate between drug- and disease-specific properties. The implementation of these concepts is illustrated by the use of clinical trial and not-in-trial simulations to complement data generation and improve benefit-risk assessment. For the sake of clarity, the proposed work is presented into three separate sections in this thesis. In **section 2**, attention is paid to importance of data quality in the context of paediatric bridging studies and the implications for the estimation of the parameters of interest in subsequent steps, i.e., evidence synthesis. Our investigation also shows how critical pharmacokinetic data are for the selection of the dosing regimen in the target population. In **section 3**, we discuss the hurdles for the assessment of efficacy in children and show that disease processes may determine the time course of response,

making drug effects no more than a covariate factor for efficacy and safety. We illustrate how treatment response can be characterised by integrating certain physiological measures (i.e., markers of pharmacology) with disease-related factors. In this context, it is also worth mentioning that further insight into the mechanisms underpinning pharmacological effects provides a systematic approach to the evaluation of safety findings. In fact, drug-disease models were developed for a series of clinically relevant outcomes, taking into account the physiological or pharmacological correlation between them. The examples presented here also provide a first insight into the concept of knowledge propagation, not as a statistical prior, but as time variant and time-invariant parameter distributions. These predictive distributions are essential in the context of chronic diseases, as they enable prediction of long-term complications or changes in response due to physiological factors as well as patient behaviour. Finally, in **section 4**, we demonstrate how MCDA can be implemented in conjunction with modelling and simulation. The models developed in the previous sections are used to generate virtual responses in clinical trial and not-in-trial simulation scenarios, mimicking a Phase III efficacy trial and a long-term follow-up pharmacovigilance protocol. The availability of a range of scenarios which have not been evaluated in an empirical setting, including predictions of long-term changes in the benefit-risk profile, provides a more robust basis for decision making regarding the approval and risk management of medicines for children.

10.1 Optimising evidence generation in paediatric trials

One of the major issues in paediatric drug development is that ethical and practical constraints often limit the generation of evidence (19,20). This has implications for the subsequent use in the evaluation of the benefit-risk profile of an intervention. In brief, there is an imperative for acquiring data with high quality and high informative value. Obviously, both the quality and informative value of data acquired in children cannot be taken for granted. Empirical experimental evidence based primarily on feasibility yields a potentially distorted picture of reality, in that drug-specific properties may not be disentangled from the role of disease-related factors and experimental design.

Given the role of extrapolation and bridging in paediatric research, in **chapters 3, 4 and 5** we demonstrate how knowledge integration can be use applied in conjunction with optimal design to evaluate which study protocol designs are more informative, whilst taking into account feasibility issues. Here we have focused on the sample size and sampling frequency required for obtaining accurate estimates of systemic exposure in children with < 6 years of age undergoing chelation therapy with deferoxamine.

The study was based on the assumption that pharmacokinetic properties can be bridged from adults and adolescents. Affected by transfusion-dependent diseases and therefore

provide evidence of the dosing regimen(s) that ensures comparable drug exposure across the overall patient population. Therefore in **chapter 3** we developed a population pharmacokinetic model using available data in adults receiving oral doses of deferiprone as a 100 mg/ml solution. Our results show how a model-based approach can be used to assess the effect of demographic and physiological factors on drug exposure and subsequently provide the basis for evaluating the design of prospective clinical trial protocols. Our analysis also illustrates how pharmacokinetic models can be used with a set of assumptions to explore the implications of factors such as co-morbidities, hepatic or renal impairment on drug exposure and consequently on dosing recommendations. In **chapter 4**, the population pharmacokinetic model describing the pharmacokinetics of deferiprone in adults and adolescents is used in conjunction with allometric scaling concepts to optimise the sampling algorithm for a prospective PK trial in children aged < 6 years. The analysis also provided an opportunity to assess the feasibility of reducing the number of patients per dose level. A sampling scheme with 5 samples post-dose per subject was found to be sufficient to ensure accurate characterization of the systemic exposure to deferiprone. Despite the assumptions regarding the changes in clearance and volume of distribution, our results reveal that the use of predefined (fixed) sampling schemes and sample sizes do not warrant accurate model structure and parameter identifiability in paediatric pharmacokinetic studies. Of importance is the accurate estimation of the magnitude of the covariate effects, as they may determine the dose recommendation for the population of interest. Furthermore, the analysis shows that the optimisation of study design does not require necessarily the use of the final model for the population of interest; the combination between ED-optimisation and the information carried by a hypothetical model is sufficient to significantly increase the quality of the information collected in a prospective clinical trial. Finally, in **chapter 5** we have performed the analysis of the pharmacokinetics of deferiprone in children aged < 6 years after administration of three different dose levels in the DEEP1 PK study (EudraCT, 2012-000658-67). The analysis also demonstrates the value of optimised protocol design, in that pharmacokinetic parameter estimates are obtained with high precision and accuracy despite sparse sampling and small sample size (i.e., 18 evaluable children with 5 samples per patients). Based on bridging concepts, a dosing regimen was recommended to this population of young children that ensures comparable exposure to adults and adolescents. An oral dose of 75 mg/kg/day deferiprone results in median AUC values of 340.6 and 318.5 $\mu\text{M}/\text{L}\cdot\text{h}$ in children and adults, respectively. Comparable values are also observed after a regimen of 100 mg/kg/day. Hence, a dosing regimen of 25 mg/kg t.i.d. should be used in children below 6 years, with the possibility of titration up to 33.3 mg/kg. The work carried out in this section allowed us to characterise the pharmacokinetics in the target population and supported the dose rationale for the subsequent assessment of the efficacy and safety of deferiprone in a non-inferiority study in the target population

From a methodological perspective, our findings highlight the role of parameter-covariate correlations to establish accurate dosing recommendations, i.e., pharmacokinetic studies in children involve more than simply generating data in a small group of children: it demands some level of stratification of the covariate factors.

10.2 Integrated evaluation of efficacy and safety by modelling and simulation

In addition to the requirement for high quality of data, accuracy and precision in the parameters of interest, the evaluation of pharmacodynamics, efficacy and safety imposes the assessment of the multidimensionality and the complexity of the clinical context in which the treatment is used. In contrast to pharmacokinetics, where measures of exposure are all derived from the underlying pharmacokinetic parameters, the analysis of pharmacodynamic data needs to account for multiple endpoints, many of which are correlated with each other. Drug-specific and system-specific parameters need to be considered in an integrated manner in order to characterise the efficacy and safety profile of a drug. As illustrated in the previous chapters of this thesis, PKPD models provide an opportunity to quantify such correlations and account for them when drawing conclusions about the benefit-risk profile of an intervention. To this end, the integration oncoming clinical data with prior knowledge (e.g. epidemiological data on background rates of expected co-morbidities; or knowledge acquired on a different disease, population or drug of the same class) becomes essential to describe the dynamics of disease and its progression and consequently determine long-term outcome.

These concepts were illustrated for characterisation of the safety and efficacy profile of deferoxamine, which is currently the first line treatment for chronic iron overload in patients affected by transfusion-dependent diseases (21–24). First, in **chapter 6** we developed a disease model for chronic iron overload based on available literature data in untreated patients. For the first time, the relationship between serum ferritin levels and blood transfusions has been characterised in a parametric manner. A turnover model was implemented in which a time-varying parameter describes the ferritin conversion rate taking into account the transfusion history and disease progression. This model provides a more mechanistic interpretation of the pathophysiological changes associated with iron overload observed during the course of transfusions. Among other things, it allows us to address some unanswered clinical questions in thalassaemia, such as to estimate the time required to achieve response based on the serum ferritin levels at the start of treatment.

This turn-over model was used as a starting point in **chapter 7** for the evaluation of the chelating effects of deferoxamine, as determined by the changes in serum ferritin levels. Deferoxamine binds iron at different extracellular levels, and within the cell it targets

lysosomal ferritin iron by stimulating ferritin degradation. The drug effect was therefore parameterised in the disease model as a proportional change in the degradation rate constant (K_{out}). Such a parameterisation can also be applied to the evaluation of other chelating agents. Most importantly, the availability of this model offers an opportunity to explore different scenarios that have been so far evaluated empirically in clinical practice. For example, it may be possible to evaluate the importance of different compliance patterns for the available chelating agents, and consequently, their impact on ferritin levels and /or risk of clinical failure. In fact, we found clear evidence that high compliance leads to stable ferritin levels over time and that poor adherence to deferoxamine therapy is strongly correlated to a poor clinical outcome

We subsequently apply this drug-disease model as a framework for further optimisation of therapeutic interventions, whereby the impact of covariate factors such as dose, drug exposure, compliance, or disease status can be evaluated against short and long-term treatment outcome.

As the assessment of the benefit-risk profile of a treatment requires quantitative descriptors of efficacy and safety, in **chapter 8**, we have complemented the work described in the previous chapter for safety endpoints. Whilst different dimensions of a symptom or sign may need to be considering when assessing its clinical relevance, here we have focused on incidence only. This decision was purely based on didactic reasons, ensuring clarity about how modelling and simulation can be used to integrate different endpoints. Two survival models were developed to describe disease-specific complications, namely hypothyroidism and type II diabetes. Both co-morbidities evolve as a consequence of iron accumulation and as such can be causally correlated with ferritin levels. A hazard function including ferritin levels was found to be a predictor of the probability of the incidence of the co-morbidity. In addition two models were developed to characterise the incidence of acute drug-specific adverse events, namely arthralgia/myalgia and anaphylaxis. They reflect two typical features of the safety profile, in that the former refers to a frequent, dose-dependent event, whereas the latter a rare, dose-independent one. Of particular relevance for the implementation of BR assessment, is the possibility of exploring rare dose-independent AEs. The four models were used in parallel to assess the impact of different exposure levels and compliance patterns on short- and long-term complications of iron chelation therapy. It should be noted that such a comprehensive analysis would not have been possible without integration of epidemiological (literature) and pharmacological data. In doing so, we have ensured that interdependencies and correlations between the different endpoints under evaluation were taken into account in a quantitative manner.

10.3 Clinical Trial Simulations: accounting for exposure, disease progression and uncertainty in benefit-risk analysis

As highlighted in the scope and intent of investigations, throughout this thesis we have defended the use of model-guided evidence generation and subsequent evidence synthesis for characterising the benefit-risk profile of medicines for children. Our results have demonstrated that empirical evidence is not necessarily accurate and that any attempt to establish the benefit-risk profile of an intervention at the time of its approval presupposes that the available data suffices to support such an assessment. This assumption may not be appropriate in a considerable number of cases. In paediatric diseases one needs to consider that the natural time course of disease occurs in parallel to developmental (physiological) growth and maturation processes. By performing clinical trial simulations and not-in-trial simulations, intrinsic and extrinsic sources of variation as well as confounding factors can be appropriately evaluated and incorporated into the decision process. The approach also addresses the issue of uncertainty due to limited sample size in clinical trials.

In **chapter 9** MCDA is used in conjunction with simulation scenarios to evaluate the benefit-risk profile deferoxamine in children with transfusion-dependent haemoglobinopathies. Here all five models developed in the previous section were used to simulate treatment response in virtual paediatric patients. Individual response data is obtained from a 1-year hypothetical phase III trial in conjunction with a follow-up safety study in which patients are evaluated up to 10 years after the start of the treatment. A reference scenario was proposed based on the currently approved dosing regimen of deferoxamine, i.e., 45 mg/kg/day (5/7). In this analysis, we have compared the results of the phase III trial with a range of alternative regimens and conditions, namely different fixed dose levels, weight-banded dosing regimens and ferritin-guided individualised regimens. The availability of simulated responses over a period of 10 years enabled us to assess the impact of long-term complications on the benefit-risk balance. Our approach clearly provides a more comprehensive evaluation of the implications of drug-specific and disease-specific factors on the overall benefit-risk profile of deferoxamine. Moreover, we show how interdependencies can be accounted for during the characterisation of long-term complications and how disease progression can be disentangled from drug-related events. The current findings open new avenues for a more structured evaluation of the BR balance of an intervention. It provides a framework for the integration of knowledge in a parametric manner, thereby 1) complementing the existing data to support the decision to be taken; 2) optimising the input data for the MCDA analysis; and 3) quantifying the relevant correlations among different endpoints and possibly determining whether personalised regimens would be of any benefit for the patient population.

10.4 Conclusions, recommendations and perspectives

Throughout this thesis we have highlighted important limitations in the assessment of BR profile of a medicinal product in children, especially if applied at the time of approval. In contrast to current practice, PKPD modelling provides a robust, mechanism-based opportunity to complement the clinical data to be used in BR assessment. Whereas different methods have been developed with the intent of enabling a more quantitative appraisal of the benefit-risk profile, none of them fully address the aforementioned limitations. Nevertheless, the MCDA appears to possess the necessary features to assess BRB in a more systematic and transparent manner, with the potential for a full integration with PKPD modelling. Yet, it should be noted that the use of MCDA has an illustrative purpose in this thesis. In principle, our approach could be implemented in combination with other quantitative BR methodologies. The major challenge lies in the steps that take place before a BR evaluation is performed. Traditional endpoints do not necessarily capture sufficient information about the treatment and the p -value of a clinical trial is not predictive of effectiveness, losing its importance in the context of BRB. This is compounded by the fact that ethical and practical constraints limit the level of clinical evidence that can be gathered in a randomised, controlled setting as well as by the effect of disease progression on the benefit-risk balance, especially in chronic conditions.

In summary, we defend the need for a development and approval paradigm in which both evidence generation and evidence synthesis form the basis for approval. Clinical events or the absence thereof are not spurious, random features of an intervention. They are greatly determined by the patient population, the context in which the treatment is assessed and by the dose rationale.

Even though some examples are available in literature where M&S is proposed in combination with clinical utility measures in the context of BR assessment (25,26), this thesis represents the first analysis in which PKPD modelling has been fully integrated with MCDA. This approach enables regulators, sponsors, and clinical experts to:

1. optimise study design, ensuring the quality of the data collected;
2. integrate available information (e.g., epidemiological data) to support data analysis and models assumptions;
3. simultaneously evaluate multiple endpoints and account for co-linearity and interdependencies and
4. most importantly, complement real data for a more comprehensive decision making.

What have we learned?

We have encountered a number of challenges that made the characterisation of treatment effects within a real-life clinical context rather complex. Currently, clinical data are generated for hypothesis testing and as such are focused on primary endpoints, not on the assessment of benefit-risk profiles. Often, the available were not sufficient to estimate all model parameters for each separate endpoint or to fully assess correlations between endpoints. Moreover, dose rather than exposure is still used as gold standard for defining treatment effects, ignoring the role of pharmacokinetics and covariate factors as explanatory variables for the variability in response.

Firstly, these challenges allowed us to learn that M&S tools provide an opportunity to describe and quantify relevant aspects of paediatric diseases even in the absence of individual data by making use of available literature as well as prior knowledge, as presented in **chapter 6**. We have shown also that despite limited evidence regarding the safety profile of deferoxamine, such a limitation does not prevent us from exploring the implications of treatment based on the integration of data from epidemiological studies as well as from a different population in which the same compound or another one of the same pharmacological/molecular class has been used. Secondly, we have shown the importance of defining a model for each endpoint to be evaluated in a BR analysis: an integrated approach with the use of multiple models is essential to characterise the multidimensionality of disease. Moreover, PKPD relationships cannot be ignored during the evaluation of the BR profile. Whereas this process was found to be resource-intensive and time-consuming, we have no doubt about its superiority in terms of establishing the true benefit-risk profile and enabling better decision making. It is also clear that implementation of the approach in a prospective manner requires efforts to be allocated as early as phase I. Finally, we have learnt that the clinical interpretation of benefit-risk estimators is fraught with a relatively large degree of uncertainty, varying considerably among the different stakeholders. These differences do not facilitate consensus regarding the consequences of an intervention. M&S allows a reduction in this uncertainty thanks to the use of underlying PKPD relationships. Such relationships are causal in nature and as such provide a somewhat more objective readout of the different criteria and their relative consequences: exposure-response data can be used to guide the expert judgment and dismiss implausible correlations. Nevertheless, we are aware of the fact that subjectivity cannot and most likely will never be fully eliminated during the appraisal of the benefit-risk profile of a medicinal product.

Requirements and recommendations

In the next paragraphs, we aim at summarising how a model-based approach can be applied to future appraisals using MCDA as a quantitative method. The first point to consider is that the clinical data generated is not sufficient for a comprehensive BR evaluation. In table 1 a

visual overview of the elements that differentiate our proposal from current practices in benefit-risk assessment. The most important message from our work is that any available knowledge on the pharmacological properties as well as on the disease and its progression cannot be omitted from a more structured and comprehensive analysis of the benefit-risk profile.

Table 1. Overview of the differences between the proposed model-based approach and the current approach for BR appraisals. CTS: Clinical Trial Simulations; NITS: Not In Trial Simulations.

CURRENT APPROACH		MODEL-BASED APPROACH
Clinical data from phase II-III trials	SOURCE	Pharmacokinetic data
		Longitudinal data
		Epidemiological data: <i>background incidences (co-morbidities and AEs)</i>
		Prior knowledge on: <i>mechanism of action; disease progression; other drugs; other populations</i>
Evidence generated	INPUT	Evidence generated + virtual scenarios (CTS and NITS*)
Tested dosing regimen vs. placebo or standard of care	OUTPUT	Alternative options: <i>possibility to achieve personalised medicine</i>

The proposed approach is versatile in that it does not necessarily rely on the characteristics of MCDA. However, if we consider M&S in the context of MCDA, as described throughout this thesis, the chart shown in figure 1 can be used to illustrate what exactly changes in benefit-risk assessment. In figure 1, the different stages of MCDA are aligned to the contributions of M&S.

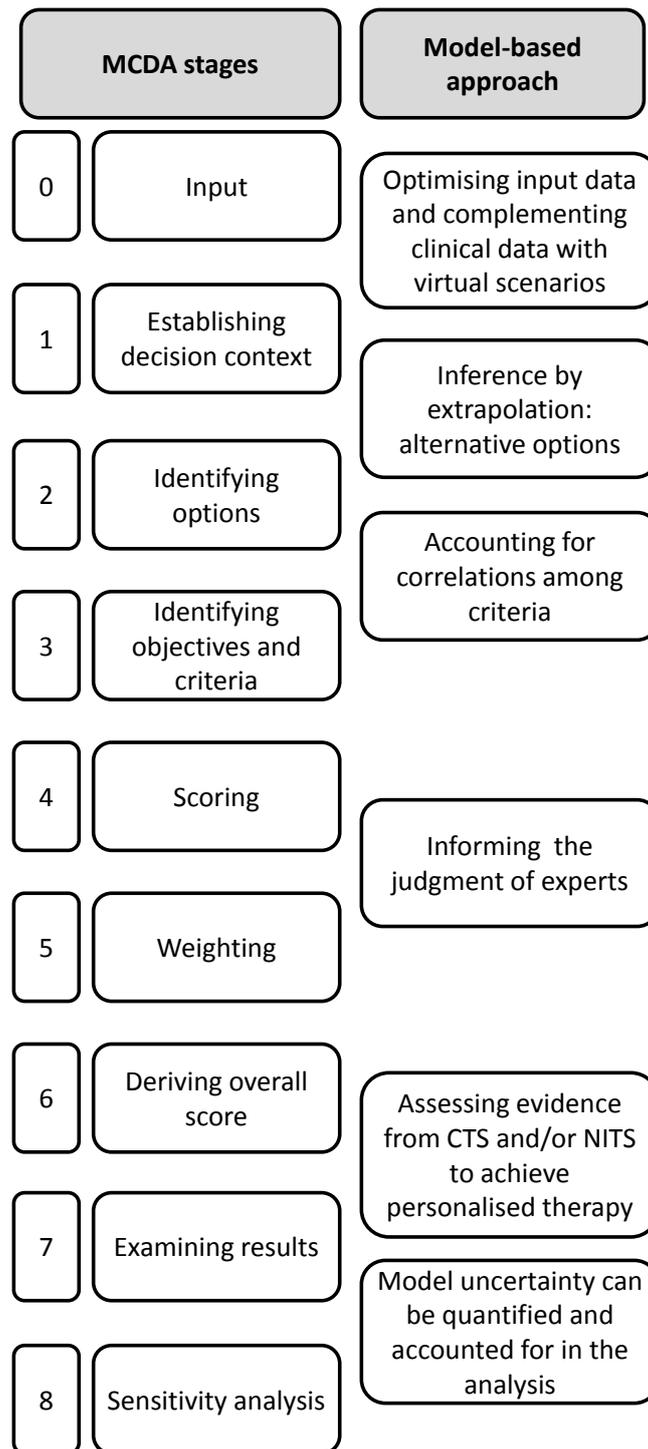


Figure 1. Contributions of the proposed model-based approach to the different stages of MCDA.

Future perspectives and conclusions

The regulation of drugs is undergoing rapid worldwide change in response to the advances in pharmaceutical sciences, drug development, and changes in public expectations. The

interest towards BR assessment is expanding and more and more projects have started focusing on the use of a more structured and transparent process by combing ideas and inputs from different stakeholders (5–8,27,28). The major effort of these groups appears to be focused on the following aspects (29–37):

1. more systematic use of available clinical evidence;
2. better graphical representation of the overall BRB;
3. re-evaluation of the BRB during the whole life cycle of the drug based on data accumulation and integration of clinical data with real data (progressive licensing);

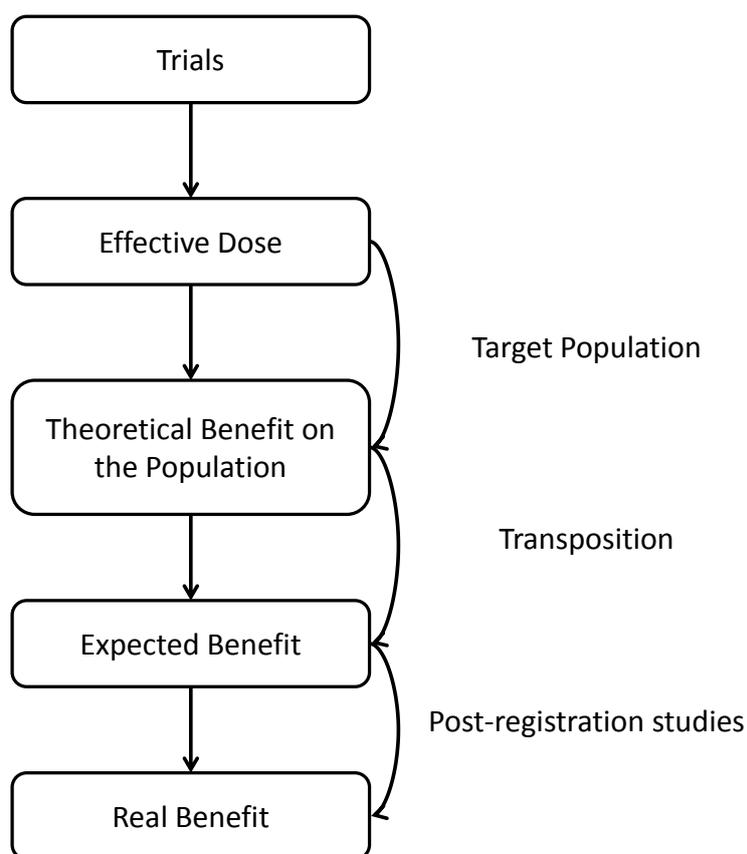


Figure 2. Process of the Public Health Benefit assessment. Adapted with permission from Massol et al. (38)

Unfortunately, as depicted in Figure 2, it appears that today's efforts rely primarily on data accumulation, making it central to the implementation of BR analysis. By contrast, we envisage the joint use of available data with drug-disease models as basis for clinical trial simulations (CTS) and/or not-in-trial simulations (NITS). The concept of extrapolating to real life population is not new and has been already applied and proposed in the context of safety management (39).

One major area that requires further development and discussion is uncertainty. While statistical uncertainty is captured well in most decision approaches, work remains to be done with regard to better articulating the consequences of any gaps in the efficacy and safety data (e.g., dropouts) and the level of evidence available on the benefit–risk profile. We acknowledge the fact that the models developed and used in this thesis carry a certain degree of uncertainty. Nevertheless, they allowed us to explore scenarios that could not be considered during drug development. They also provided answers to clinical questions (e.g., impact of long-term complications on the BRB of iron chelators) that could not necessarily be addressed directly in a real setting. Drug development and therapeutics will greatly benefit from a framework that describes how drug- and disease-specific properties interact with each other and ultimately determine the benefit-risk profile during development (i.e. randomised clinical trials) as well as during clinical use of the drug.

Our approach could form the backbone for the recently proposed progressive licensing model, which was initiated by Health Canada to develop a drug regulatory system for the future (36). The progressive licensing model consists in sound scientific evidence and risk management. It is aimed at supporting access to promising new drugs and the continuous monitoring of safety, quality, and efficacy. It is being developed on the assumption that knowledge and experience can be gained from every stage of a drug's life cycle. A well-designed regulatory framework should support the collection, analysis, and communication of knowledge and experience about a drug throughout its life cycle so that it can be used wisely. In addition, in contrast to network meta-analysis which relies in stochastic parameterisation of the trade-offs between risk and benefit, the use of drug disease models suits the same purpose using a biologically, clinically plausible parameterisation (40).

In conclusion, it should be highlighted that models do not make decisions, people do. A collaborative effort between industry and regulators will be required to continue to advance the science of benefit–risk methodology, since, as we have argued above, there is no single or simple approach that would address all benefit–risk assessments. Eventually, we expect a set of common principles, standards and a toolbox of methods will emerge. Ultimately, patients, clinicians, drug developers and regulators need to acknowledge that decisions are better made when data are presented and communicated in a clear, systematic manner. PKPD modelling can complement evidence generation by providing stakeholders the opportunity to explore conditions that have not been experimentally tested at the time of BR analysis. Regardless of the limitations models and simulation scenarios may have, model-based evaluation is likely to outperform gut feeling, which currently prevails in clinical decision-making.

References

1. Downing NS, Aminawung J a, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA*. 2014;311(4):368–77.
2. Food and Drug Administration (FDA). Guidance for Industry: Exposure-Response Relationships - Study Design, Data Analysis and Regulatory Applications. 2003;(April):1–25.
3. Food and Drug Administration (FDA). Guidance for Industry Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products. 1998;(May):1–23.
4. Katz R. FDA: evidentiary standards for drug development and approval. *NeuroRx*. 2004;1(July):307–16.
5. Arnardóttir AH. Regulatory benefit-risk assessment. Different perspectives. 2012.
6. Levitan B, Mussen F. Evaluating benefit–risk during and beyond drug development: an industry view. *Regul Rapp Focus – Benefit–risk Assess Fram*. 2012;9(6):10–5.
7. Food and Drug Administration (FDA). Structured Approach to Benefit-Risk Assessment in Drug Regulatory Decision-Making. 2013.
8. Mt-Isa S, Wang N, Hallgreen CE, Callréus T, Genov G, Hirsch I, et al. PROTECT - Review of methodologies for benefit and risk assessment of medication. 2012.
9. Coplan P, Noel R, Levitan B, Ferguson J, Mussen F. Development of a Framework for Enhancing the Transparency , Reproducibility and Communication of the Benefit – Risk Balance of Medicines. *Clin Pharmacol Ther*. 2011;89:312–5.
10. EMA. Benefit-risk methodology project - Project description. 2009.
11. FDA. Structured Approach to Benefit-Risk Assessment in Drug Regulatory Decision-Making - Draft PDUFA V Implementation Plan. 2013.
12. Liberti B, McAuslane N, Walker S. Progress on the Development of a Benefit / Risk Framework for Evaluating Medicines. *Regul Focus*. 2010;15:32–7.
13. Holden WL. Benefit-risk analysis : a brief review and proposed quantitative approaches. *Drug Saf*. 2003 Jan;26(12):853–62.
14. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. *Value Health*. 2010 Aug;13(5):657–66.
15. Walker S, McAuslane N, Liberti L, Salek S. Measuring benefit and balancing risk: strategies for the benefit-risk assessment of new medicines in a risk-averse environment. *Clin Pharmacol Ther*. 2009;85:241–6.

16. Ariens E, Van Rossum J, Simonis A. Affinity, intrinsic activity and drug interactions. *Pharmacol Rev.* 1957;9(2):218–36.
17. Patil P, Everhardus J. Ariens (1918-2002): a tribute. *Trends Pharmacol Sci.* 2002;23(7):344–5.
18. Timmerman H, Breimer D. Everhardus Jacobus Ariens. *Sciences TRNA of A and*, editor.
19. ICH Topic E11.
http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002926.pdf. 2001.
20. Sammons HM, Starkey ES. Ethical issues of clinical trials in children. *Paediatr Child Health (Oxford)*. Elsevier Ltd; 2012 Feb;22(2):47–50.
21. Fisher S, Brunskill S, Doree C, Gooding S, Chowdhury O, Roberts D. Desferrioxamine mesylate for managing transfusional iron overload in people with transfusion-dependent thalassaemia (Review). *cochrane Libr.* 2013;(8).
22. Galanello R, Origa R. Beta-thalassemia. *Orphanet J Rare Dis.* 2010 Jan;5:11.
23. Rebullia P. Blood transfusion in beta thalassaemia major. *Transfus Med.* 1995 Dec;5(4):247–58.
24. Rund D, Rachmilewitz E. Beta-Thalassemia. *N Engl J Med.* 2005;353:1135–46.
25. Ouellet D. Benefit-risk assessment: the use of clinical utility index. *Expert Opin Drug Saf.* 2010 Mar;9(2):289–300.
26. Ouellet D, Werth J, Parekh N, Feltner D, McCarthy B, Lalonde RL. The use of a clinical utility index to compare insomnia compounds: a quantitative basis for benefit-risk assessment. *Clin Pharmacol Ther.* 2009 Mar;85(3):277–82.
27. European Medicines Agency (EMA), Committee for medicinal products for human use (CHMP). Reflection paper on benefit-risk assessment methods in the context of the evaluation of marketing authorisation applications of medicinal products for human use. 2008.
28. Curtin F, Schulz P. Assessing the benefit:risk ratio of a drug - randomized naturalistic evidence. *Dialogues Clin Neurosci.* 2011;13:183–90.
29. Hallgreen CE, van den Ham HA, Mt-Isa S, Ashworth S, Hermann R, Hobbiger S, et al. Benefit–risk assessment in a post-market setting: a case study integrating real-life experience into benefit–risk methodology. *Pharmacoepidemiol Drug Saf.* 2014;23(July):974–83.
30. Hammad TA, Neyrapally GA, Iyasu S, Staffa JA, Dal Pan G. The Future of Population-Based Postmarket Drug Risk Assessment: A Regulator’s Perspective. *Clin Pharmacol Ther.* 2013;94(3):349–58.

31. Sarac SB, Rasmussen CH, Rasmussen M a, Hallgreen CE, Sjøeborg T, Colding-jørgensen M, et al. A Comprehensive Approach to Benefit-Risk Assessment in Drug Development. *Basic Clin Pharmacol Toxicol*. 2012;111:65–72.
32. Verhagen H, Tijhuis MJ, Gunnlaugsdóttir H, Kalogeras N, Leino O, Luteijn JM, et al. State of the art in benefit–risk analysis: Introduction. *Food Chem Toxicol*. 2012;50:2–4.
33. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines – part 1 : The development of a new model using multi-criteria decision analysis. *Pharmacoepidemiol Drug Saf*. 2007;(June):2–15.
34. Ashby D. *Visualizing Benefit-Risk for Drug Regulations*. 2014.
35. Dal Pan GJ. *Benefit-risk balance assessment during life cycle of medicinal products*. 2013.
36. Lourenco C. *Concept of Benefit / Risk Presentation to APEC Preliminary Workshop on Review of Drug Development in Clinical Trials*. 2008.
37. Rasmussen CH, Sarac SB. *Benefit-Risk Assessment from a Clinical Point of View: a structured approach with focus on transparency, clinical significance and visualization*. 2012.
38. Massol J, Puech A, Boissel J. *How to Anticipate the Assessment of the Public Health Benefit of New Medicines?* *Therapie*. 2007;62(5):427–35.
39. Chain ASY, Dieleman JP, van Noord, Charlotte Hofman A, Stricker BHC, Danhof M, Sturkenboom MCJM, et al. *Not-in-trial simulation I: Bridging cardiovascular risk from clinical trials to real-life conditions*. *Br J Clin Pharmacol*. 2013;76(6):964–72.
40. Van Valkenhoef G, Tervonen T, Zhao J, De Brock B, Hillege HL, Postmus D. *Multicriteria benefit-risk assessment using network meta-analysis*. *J Clin Epidemiol*. Elsevier Inc; 2012;65(4):394–403.

CHAPTER 11

Nederlandse samenvatting

(Synopsis in Dutch)

De evaluatie en de toelating van geneesmiddelen voor kinderen is voortdurend aan verandering onderhevig. Enerzijds is dat een gevolg van het toegenomen besef van het belang de effectiviteit en veiligheid van geneesmiddelen voor deze doelgroep zo nauwkeurig mogelijk vast te stellen. Anderzijds is het een gevolg van vooruitgang in statistische en klinische methodologieën die het mogelijk maakt het baten-risico profiel steeds nauwkeuriger vast te stellen. Omdat het moeilijk is om voldoende onderzoeksresultaten te verkrijgen voor de toepassing van geneesmiddelen voor kinderen, wordt vaak gebruik gemaakt van gegevens verkregen nadat een geneesmiddel is toegelaten (de zogenoemde '*post-approval data*'). Hierbij geldt een *fait accompli*, oftewel het bewijs is verzameld na de feiten.

Op dit ogenblik is er geen kwantitatief raamwerk voor autoriteiten, klinische onderzoekers en bedrijven die geneesmiddelen ontwikkelen, op basis waarvan de bestaande kennis en informatie over zowel het geneesmiddel als de ziekte kan worden geïntegreerd om daarmee de effectiviteit van het geneesmiddel te voorspellen voordat het onderzoek wordt gestart. Door een kwantitatief raamwerk op te stellen is het niet alleen mogelijk om het risico management plan te optimaliseren, het geeft ook houvast om tijdens het geneesmiddel ontwikkelingsproces wetenschappelijke en klinische vragen te beantwoorden.

Het onderzoek beschreven in dit proefschrift is er op gericht om op basis van kwantitatieve klinische farmacologische principes aanvullend bewijs te verkrijgen over de werkzaamheid van geneesmiddelen bij kinderen. Hiervoor is modelering en simulatie van klinische farmacokinetische en farmacodynamische gegevens gebruikt. In hoofdstuk 1 laten we zien hoe voorspellingen op basis van een modelmatige aanpak kunnen worden gecombineerd met bestaande methodes voor het karakteriseren van het baten en risico profiel van een geneesmiddel. Op basis daarvan verwachten we het besluitvormingsproces voor de toelating van geneesmiddelen voor kinderen te verbeteren. Twee belangrijke zaken onderscheiden het onderzoek beschreven in dit proefschrift van eerder onderzoek in pediatrie klinische farmacologie. Wij hebben voor het eerst meerdere geneesmiddel-ziekte modellen tegelijkertijd geanalyseerd en gesimuleerd waarbij rekening werd gehouden met mogelijke correlaties tussen effectiviteit en veiligheid. Dit levert een belangrijke bijdrage aan de manier waarop het effect van een geneesmiddel wordt geanalyseerd; niet zozeer als primair eindpunt in een klinische studie, maar als middel om geneesmiddeleigenschappen te onderscheiden van ziekte eigenschappen. Hierdoor wordt het mogelijk om de effectiviteit en veiligheid van

een geneesmiddelinterventie te beschrijven met behulp van parameters. Het tweede belangrijke punt waar ons onderzoek aan heeft bijgedragen, is het gebruik van simulaties van effecten in klinische studies en in patiënten die gewoonlijk worden uitgesloten van deelname aan klinische studies, de zogenoemde “not-in-trial” populatie, als toevoeging aan de klinische gegevens. Hierbij worden gesimuleerde gegevens uit virtuele scenarios verweven met echte gegevens en vervolgens gebruikt als input voor MCDA (Multi Criteria Decision Analysis). MCDA wordt gezien als hulpmiddel bij het oplossen van vraagstukken waarbij de oplossing moet voldoen aan meerdere, mogelijks conflicterende eisen. Het voordeel van deze aanpak is dat het mogelijk is om al voor de beoordeling en toelating van een geneesmiddel, een kwantitatieve analyse te doen van het baten-risico profiel in situaties die niet getest zijn tijdens de ontwikkelingsfase(s). Dit is vooral van belang voor de evaluatie van geneesmiddelen voor kinderen, omdat er maar op zeer beperkte schaal gegevens verkregen kunnen worden uit deze populatie. Daarnaast kan men rekening houden met de rol van fysiologische processen zoals rijping en groei die op lange termijn de baten-risico balans van de behandeling kunnen beïnvloeden.

De punten zoals zojuist besproken komen in de verschillende hoofdstukken van dit proefschrift aan de orde. Het voorgestelde raamwerk voor de toepassing van een modelmatige aanpak bij de toelatingsprocedure van geneesmiddelen voor kinderen wordt geïllustreerd aan de hand van de ontwikkeling van chelatietherapie voor ijzerstapelingsziekten als gevolg van herhaalde bloedtransfusie.

11.1 Chronische ijzerstapeling door bloedtransfusie-afhankelijke hemoglobinopathien

Bèta-thalassemie-major is één van de meest voorkomende bloedtransfusie-afhankelijke ziekten. Het is een erfelijke ziekte die door een sterk gereduceerd of volledig afwezige synthese van bèta-globine wordt veroorzaakt. Hierdoor wordt in het lichaam van patiënten met bèta-thalassemie-major onvoldoende en afwijkend hemoglobine (Hb) in de rode bloed cellen (RBC) aanmaakt. Ook daalt de productie van RBC waardoor deze patiënten bloedarmoede hebben. Daarom hebben patiënten met bèta-thalassemie-major regelmatig een bloedtransfusie nodig om te kunnen overleven. Alhoewel de behandeling met chronische bloedtransfusie sterk verbeterd is in de afgelopen jaren, zijn er nog steeds een groot aantal complicaties. Door de vele bloedtransfusies kan ijzerstapeling optreden en dit kan hartfalen, lever fibrose, suikerziekte en een verstoorde hormoon productie tot gevolg hebben. Omdat het menselijk lichaam zelf geen mechanisme heeft om een overschot aan ijzer af te voeren is een adequate ontijzering noodzakelijk om complicaties te voorkomen. Hiervoor worden ijzerchelatoren gebruikt die 1) voorkomen dat niet aan transferrine gebonden ijzer (NTBI)

wordt opgenomen door organen zoals hart en lever, 2) voorkomen dat intracellulair ijzer wordt opgenomen door ferritine, en 3) ijzer wegvangen van ferritine dat wordt afgebroken. Aangezien de rol van ferritine bij ijzerstofwisseling, wordt de ferritine concentratie in serum gebruikt als maat voor ijzerstapeling. Daarvoor geldt de aanname dat ferritine aan de totale ijzerhoeveelheid in het lichaam gecorreleerd kan worden. Het bepalen van de serum ferritine concentratie is een eenvoudige en minimaal invasieve methode om de ijzerstapeling en het beloop van het effect van chelatietherapie te volgen.

11.2 Optimalisatie van het verzamelen van gegevens bij geneesmiddelonderzoek in kinderen

Klinisch onderzoek in kinderen is beperkt door zowel ethische als praktische oorzaken. Daarom worden extrapolatie en overbruggingsconcepten vaak gebruikt als basis voor de evaluatie van geneesmiddelen in paediatrische indicaties. Dit maakt het lastig om het batenrisico profiel van een therapie op te stellen. Daarnaast moet men ook erop letten dat de gegevens die worden verzameld van voldoende kwaliteit zijn.

Rekening houdend met het bovenstaande hebben we in **hoofdstukken 3, 4 en 5** beschreven hoe een klinisch studieprotocol de meeste informatie kan opleveren. Hiervoor hebben we zowel de mogelijkheden van integratie van kennis als de praktische uitvoerbaarheid in overweging genomen. Onze belangrijkste doelstellingen waren om te onderzoeken hoe veel patiënten in een studie nodig zijn en hoe vaak bloedmonsters afgenomen moeten worden om de blootstelling aan een geneesmiddel nauwkeurig te bepalen. Het optimalisatie concept is geïllustreerd door het vaststellen van de farmacokinetische eigenschappen van deferiprone in kinderen jonger dan 6 jaar. Door aan te nemen dat de farmacokinetiek van deferiprone kan worden voorspeld op basis van gegevens verkregen uit klinisch onderzoek in volwassenen en in jongvolwassenen met bloedtransfusie-afhankelijke ziekten, kan het doseringsschema en dus de blootstelling gelijk worden gehouden voor de totale patiënten populatie. Op basis van beschikbare gegevens in volwassenen die werden behandeld met een orale dosis van 75 mg/kg/day deferiprone hebben we in **hoofdstuk 3** een populatie farmacokinetisch model ontwikkeld. Onze resultaten laten zien dat een modelmatige aanpak gebruikt kan worden om het effect van demografische en fysiologische factoren op het lotgeval van deferiprone te karakteriseren. Kennis omtrent de farmacokinetiek van deferiprone kan vervolgens geïntegreerd worden met statistische beginselen om nieuwe klinische studies te ontwerpen. De parameters die de farmacokinetiek van deferiprone beschrijven zijn in **hoofdstuk 4** toegepast in combinatie met allometrische schaling om de optimale tijdstippen voor bloedafname te bepalen. Hieruit bleek dat het voldoende is om per patiënt 5 bloedmonsters af te nemen om de systemische blootstelling aan deferiprone nauwkeurig vast te stellen in kinderen. Ook liet deze analyse zien dat het niet noodzakelijk is om voor de optimalisatie van

de opzet van een klinische studie een definitief farmacokinetisch model ter beschikking te hebben. De toepassing van geëxtrapoleerde parameters bij optimalisatieprocedures in paediatrisch onderzoek is een aanzienlijk nieuw concept. In **hoofdstuk 5** hebben we, uit data verkregen uit de DEEP1 PK studie (EurdrACT, 2012-000658-67), de farmacokinetiek van deferiprone in kinderen jonger dan 6 jaar geanalyseerd. Hieruit bleek de meerwaarde van een geoptimaliseerd protocol ontwerp erg duidelijk: de farmacokinetische parameters konden met grote precisie worden bepaald ondanks het kleine aantal patiënten (18 kinderen) dat in de studie was opgenomen en het kleine aantal bloedmonsters (5 per patiënt). Op basis van overbruggingsconcepten hebben we een doseringsschema voor deferiprone in kinderen jonger dan 6 jaar kunnen vaststellen zodat de blootstelling aan deferiprone bij deze groep patiënten vergelijkbaar is met die van volwassenen en jongvolwassenen. Samenvattend tonen deze resultaten aan, dat modelmatig ontworpen overbruggingsprotocollen robust zijn om de farmacokinetiek in een nieuwe doelgroep te karakteriseren. Bovendien, kunnen deze resultaten de basis vormen voor het kiezen van de dosis en doseerschema voor het vervolgonderzoek waarin de effectiviteit en veiligheid van deferiprone in de doelgroep wordt bepaald. Vanuit een methodologisch perspectief blijkt het dat de correlatie tussen farmacokinetische parameters en covariaten bepalend zijn om een nauwkeurige aanbeveling over de juiste dosering te kunnen maken. Oftewel: farmacokinetische studies in kinderen is meer dan data verzamelen in een kleine studiegroep, het identificeren van de covariaten op basis waarvan de variabiliteit in farmacokinetische parameters kan worden bepaald speelt hierbij een grote rol.

11.3 Geïntegreerde evaluatie van effectiviteit en veiligheid door modelering en simulatie

Voor het bepalen van de effectiviteit en veiligheid van een geneesmiddel is meer nodig dan een betrouwbare meting en een precieze bepaling van relevante eindpunten en parameters. Daarvoor dienen de complexiteit van de ziekte en de onderliggende klinische context ook meegenomen te worden. In tegenstelling tot een farmacokinetische analyse waarbij de mate van blootstelling direct uit enkele parameters wordt bepaald, zijn er bij het bepalen van de farmacodynamiek vaak meerdere eindpunten nodig, die ook onderling een samenhang met elkaar hebben. Om de effectiviteit en veiligheid van een geneesmiddel te kunnen vaststellen moet een onderscheid worden gemaakt tussen de geneesmiddel-specifieke parameters en de ziekte-specifieke parameters. In de voorafgaande hoofdstukken van dit proefschrift hebben we laten zien dat PKPD modellen gebruikt kunnen worden om de correlatie tussen de geneesmiddel- en ziekte-specifieke parameters te karakteriseren. Daardoor is het ook mogelijk om met deze correlatie rekening te houden wanneer het baten-risico profiel van een geneesmiddel wordt opgesteld.

Als men de lange termijn uitkomsten van een behandeling tracht te voorspellen is het essentieel om nieuwe onderzoeksgegevens over de dynamiek en het verloop van een ziekte te integreren met bestaande kennis. De bestaande kennis kan worden verkregen of afgeleid uit oa. epidemiologische gegevens over verwachte co-morbiditeit(en), uit andere ziektemodellen of patiëntengroepen, of uit geneesmiddelen uit dezelfde klasse.

Om deze concepten te illustreren hebben we de effectiviteit en veiligheid van deferoxamine gekarakteriseerd. Deferoxamine is een ijzerchelator die op dit moment als eerstelijns therapie bij chronische ijzerstapeling wordt voorgeschreven. In eerste instantie hebben we in **hoofdstuk 6** een ziektemodel voor chronische ijzerstapeling ontwikkeld op basis van literatuurgegevens die de ijzer homeostase in onbehandelde patiënten beschrijft. In dit “turnover” model, hebben we gebruik gemaakt van een tijdsafhankelijke parameter die de omzettingssnelheid van ferritine weergeeft waarbij rekening werd gehouden met bloedtransfusies en ziekte progressie. Dit model sluit een mechanistische benadering om van de pathofysiologische veranderingen als gevolg van ijzerstapeling door herhaalde bloedtransfusie. In **hoofdstuk 7** is dit ziektemodel gebruikt om op basis van van veranderingen in serum ferritine concentraties, het effect van deferoxamine te bepalen. Daardoor kunnen verschillende scenario's gesimuleerd worden waar dit tot nu toe slechts empirisch konden worden bepaald. Het is bijvoorbeeld mogelijk om de invloed van patiënt-trouw voor de verschillende ijzerchelatoren te simuleren en daarbij het effect op de ferritine concentratie in serum te bepalen. Daaruit kan men oa de klinische respons en het succes of faal van de behandeling te voorspellen. Dit geneesmiddel-ziekte model hebben we vervolgens toegepast als raamwerk voor de optimalisatie van de respons op ijzerchelatoren. Hierbij kunnen covariaten zoals dosering, blootstelling, patiënt-trouw, stadium van de ziekte worden afgezet tegen korte- en lange termijn uitkomsten.

Omdat het voor het bepalen van het baten-risico profiel van een geneesmiddel van belang is om niet alleen de werkzaamheid te bepalen, hebben we in **hoofdstuk 8** ook de veiligheid (bijwerkingen en ziektecomplicaties) ook door middel van parametrische methodes samengevat. Hiervoor hebben we twee overlevings-modellen ontwikkeld om hypothyroidie en diabetes mellitus als ziekte-specifieke complicaties te kunnen beschrijven. Zowel diabetes als hypothyroidie zijn een co-morbiditeit van ijzerstapeling en kunnen daarom worden gecorreleerd aan de ferritine concentratie in serum. Een risico functie waarbij rekening gehouden werd met de ferritine concentratie bleek het ontstaan van de co-morbiditeiten te kunnen voorspellen. Daarnaast hebben we twee modellen ontwikkeld om de incidentie van gewrichts- en spierpijn en overgevoelighedsreacties, twee acute en geneesmiddel-specifieke bijwerkingen van ijzerchelatoren, te kunnen karakteriseren. Deze vier modellen werden tevens gebruikt om de invloed van blootstelling en patiënt-trouw op zowel de korte- als de

lange termijn complicaties van ijzerchelatie therapie te onderzoeken. Hierbij dient te worden opgemerkt dat deze uitvoerige analyse alleen mogelijk was door de integratie van gegevens uit epidemiologische en klinisch farmacologische studies. Hierdoor hebben we zowel de correlatie tussen als de tijdsafhankelijkheid van de verschillende eindpunten op een kwantitatieve manier kunnen meewegen in onze analyse.

11.4 Simulatie van klinische studies als basis voor de baten-risico analyse

Het doel van ons onderzoek is om de meerwaarde van modelering en simulatie voor het genereren van zo een realistisch mogelijk baten-risico profiel voor geneesmiddelen voor kinderen. Met het onderzoek beschreven in dit proefschrift willen we aantonen dat het bewijs dat empirisch verkregen wordt niet altijd accuraat is. En dat de aanname dat de gegevens die beschikbaar zijn op het moment van goedkeuring voldoende zijn om een baten-risico profiel voor een nieuw geneesmiddel op te stellen, ook niet altijd klopt. Door gebruik te maken van simulatie van de effecten verkregen uit klinische studies en in individuen die gewoonlijk worden uitgesloten van deelname aan klinische studies (“not-in-trial” simulaties), is het mogelijk om rekening te houden met intrinsieke en externe variabiliteit, zoals bijvoorbeeld verschillen in blootstelling en ziekte progressie. Ook is het mogelijk om eventueel versturende factoren te analyseren zodat hiermee rekening gehouden kan worden in het besluitvormende proces. Deze aanpak maakt het ook mogelijk om onzekerheden door het kleine aantal patiënten (en/of bloedmonsters) in de klinische studies uit te sluiten.

In **hoofdstuk 9** hebben we het baten-risico profiel voor deferoxamine bij kinderen met een bloedtransfusie-afhankelijke hemoglobinopathie geëvalueerd door gebruik te maken van MCDA technieken in combinatie met gesimuleerde scenario's, waarbij virtuele patienten op verschillende manieren worden behandeld. In onze analyse hebben we de resultaten van een standaard fase III studie vergeleken met de gesimuleerde resultaten van verschillende behandeling condities en doseringsschema's voor deferoxamine. Op deze manier konden we de gevolgen van lange-termijn complicaties op het baten-risico profiel kwantificeren. Deze resultaten laten zien dat het mogelijk is om het baten-risico profiel van een geneesmiddel op een meer gestructureerde manier te evalueren voordat data omtrent de effectiviteit van de behandeling wordt verkregen. Bovendien laat ons onderzoek een raamwerk zien waarbij bestaande kennis over de ziekte, patiënten populatie en het geneesmiddel wordt geïntegreerd als parametrische verdelingen in een groep modellen. Dit heeft een drietal voordelen: 1) bestaande gegevens worden aangevuld om zo de besluitvorming te verbeteren; 2) de input voor de MCDA wordt geoptimaliseerd; en 3) dankzij het vaststellen van relevante correlaties tussen verschillende eindpunten wordt het mogelijk om te evalueren of een

bepaalde dosis of doseringsschema een voordeel heeft voor de gehele patiënten populatie of alleen maar voor een deel daarvan. Op basis van zulke scenario's kan besloten worden of een gepersonaliseerde therapie zinvol, wenselijk of onnodig is voor de betrokken patiënten.

11.5 Discussie, conclusies en aanbevelingen

In dit proefschrift hebben we laten zien dat de huidige manier om een baten-risico profiel voor een geneesmiddel voor kinderen op te stellen niet afdoende is. Vooral niet als men zich realiseert dat de evaluatie van baten en risico's zich beperkt tot gegevens die beschikbaar zijn op het moment van de toelating van het desbetreffende geneesmiddel. Ondanks de beschikbare methoden om het baten-risico profiel op een kwantitatieve manier te analyseren, hebben ze een belangrijke tekortkoming, namelijk alle methoden zijn puur data gedreven. Als zodanig blijken ze ongeschikt te zijn om baten en risico's te voorspellen voordat alle relevante gegevens verkregen worden. Door gebruik te maken van farmacokinetische en farmacodynamische modelering, een robuuste en mechanistische methode waardoor ziekte- en geneesmiddel-specifiek eigenschappen worden meegewogen, kunnen bestaande klinische gegevens aangevuld en versterkt worden. Aangezien MCDA technieken de besluitvorming uiteindelijk tot getallen en cijfers omzetten, biedt deze aanpak de mogelijkheid om een baten-risico profiel op een systematische en transparante manier te analyseren. Een bijkomend voordeel is dat MCDA geïntegreerd kan worden met farmacokinetische en farmacodynamische modelering. Hierdoor kan een betrouwbaarder baten-risico profiel opgesteld worden, rekening houdende met de invloed van zowel het beloop van de ziekte als het gebrek aan experimentele data.

In dit proefschrift werd de MCDA gekozen om de geïntegreerde benadering te implementeren. Dit concept kan echter ook worden gebruikt in combinatie met andere kwantitatieve baten-risico methoden. In feit, er zijn verschillende voorbeelden in de wetenschappelijke literatuur waarbij modelering en simulatie gebruikt wordt in combinatie met een klinische utiliteitsfuncties of waarderingsfuncties. In tegenstelling tot die voorbeelden, is in dit proefschrift voor het eerst gebruik gemaakt van een methode waarbij PKPD modelering volledig geïntegreerd is met MCDA. Aangezien het besluitvormingsproces aanzienlijk verbeterd kan worden biedt deze aanpak belangrijke voordelen voor regulatoire autoriteiten, farmaceutische bedrijven en klinische deskundigen:

- 1) Het studie protocol kan worden geoptimaliseerd. Hierbij wordt de kwaliteit van de data verbeterd en gewaarborgd voordat patiënten geïncludeerd worden in een klinische studie;
- 2) Beschikbare kennis (bijvoorbeeld epidemiologische gegevens) kan op een formele manier worden geïntegreerd om de data analyse en model-aannames te versterken;

- 3) Meerdere eindpunten kunnen simultaan worden geëvalueerd waarbij rekening wordt gehouden met onderlinge correlaties en tijdsafhankelijkheid; en
- 4) Gegevens die verkregen zijn tijdens de klinische ontwikkelingsfase kunnen worden aangevuld met gesimuleerde data, waardoor de onzekerheden omtrent de baten en risico's meegewogen worden tijdens de analyse.

Het vermogen tot extrapolatie stelt de analist of onderzoeker in staat nieuwe alternatieven te evalueren en te klasseren, en dit alles in *real-time*. Alhoewel onze aanpak niet per se gekoppeld hoeft te worden aan MCDA, zijn er een aantal belangrijke veranderingen nodig om de technieken met elkaar te kunnen integreren (Tabel 1).

Tabel 1. Bijdrage van de voorgestelde modelmatige aanpak aan de verschillende stappen van MCDA.

Multi-criteria decision analysis	Modelmatige aanpak
0: input	
1: bepalen van besluitvorming	0-1: Optimalisatie van input gegevens en aanvullen van klinische gegevens met virtuele scenario's
2: identificeren van uiteenlopende opties	1-2: gevolgtrekking door extrapolatie: alternatieve opties
3: identificeren van doelstelling en criteria	2-3: rekening houden met correlaties tussen criteria
4: scoring	
5: weegfactoren	4-5: besluitvormingssleutels en waarderingfuncties door experts
6: bepalen van overall score	
7: analyseren van resultaten	6-7: beoordelen van bewijs van CTS en/of NITS om gepersonaliseerde geneesmiddeltherapie te bereiken
8: gevoeligheidsanalyse	7-8: onzekerheden in het model kunnen worden gekwantificeerd en er kan rekening mee worden gehouden in de analyse

Toekomstperspectief en conclusie

De wetgeving rondom de goedkeuring van geneesmiddelen is wereldwijd aan verandering onderhevig. Dit komt niet alleen door vooruitgang in de farmaceutische wetenschappen en geneesmiddelontwikkeling, maar ook door veranderingen in het verwachtingspatroon van patiënten. Er is een toegenomen belangstelling voor het opstellen van een baten-risico profiel voor geneesmiddelen. Steeds meer onderzoeksprojecten maken gebruik van een gestructureerd en transparant proces waarbij rekening wordt gehouden met de eisen en

verwachtingen van verschillende belanghebbenden. De belangrijkste punten die hierbij consequent naar voren komen zijn:

- 1) Systematisch gebruik van beschikbaar klinisch bewijs;
- 2) Verbeterde grafische weergave van het totale baten-risico profiel;
- 3) Een her-evaluatie van het baten-risico profiel gedurende de levensloop van een geneesmiddel door nieuwe (post-market) gegevens te integreren met data uit de klinische studies, oftewel adaptive licensing.

Vanuit een theoretisch perspectief vergen de huidige methoden het verzamelen van nieuwe gegevens om baten en risico's te kunnen beoordelen (Table 2) . Wij willen juist stimuleren dat het simultaan gebruik van beschikbare gegevens met geneesmiddel-ziekte modellen de basis kunnen zijn voor simulatie van klinische studies en/of 'not in trial'simulaties.

Tabel 2. Overzicht van de verschillen tussen de huidige aanpak en de voorgestelde modelmatige aanpak voor baten-risico analyse. CTS: Clinical Trial Simulations; NITS: Not-In-Trial Simulations.

Huidige aanpak		Modelmatige aanpak
Klinische data van fase II-III studies	Bron	Farmacokinetische data
		Longitudinale data
		Epidemiologische data: <i>achtergrond frequentie (co-morbiditeiten en bijwerkingen)</i>
		Achtergrond kennis van: <i>werkingsmechanisme; ziekte progressie; andere geneesmiddelen; andere doelgroepen</i>
Op basis van bewijs uit gecontroleerde klinische studies	INPUT	Op basis van klinische data én virtuele patiënten (CTS en NTIS)
Aanbevelingen vaak beperkt tot de beschikbare data, namelijk de geteste doseringsschema's en/of patiënten populaties	OUTPUT	Aanbevelingen ook voor gextrapoleerde scenario's en/of andere patiënten populaties: <i>ontwikkelen van gepersonaliseerde geneesmiddeltherapie</i>

Acknowledgments

The past few years that I spent at the Division of Pharmacology in Leiden happened to overlap with a period of substantial turnaround of staff and students. This gave me the chance to get to know quite a large group of people with whom I have not only shared my time at the department but most importantly also ideas and experience. In a way or another, this interaction led them undoubtedly to contribute to the work described in this thesis. As the list would probably fill this whole page, so I prefer to simply recognise everyone's contribution with a big Thank You!

I would like to make a special reference to my supervisors, Oscar and Meindert, for the brilliant ideas and the guidance they have provided me with during this challenging journey. I am also thankful to three students I had the privilege to supervise: Annette, Morris and Rob; they have contributed with great enthusiasm and effort to the work described in chapters 1, 6 and 8, which provided the basis for the development of almost all the concepts of this thesis. Last but not least, I would like to thank my paranimphs Chiara and Sven for always being there. I am grateful for having met you guys, you have been during this whole period and still are not only a huge support but above all very good friends.

Se oggi ho avuto la possibilità di raggiungere anche questo traguardo lo devo soprattutto alle persone che mi stanno vicino e che mi hanno permesso di crescere ogni giorno: la mia Famiglia. Grazie!

Curriculum vitae

Francesco Bellanti was born in Messina, Italy, on the 10th of January 1986. He obtained his scientific diploma in 2004 at the Liceo Scientifico “G. Seguenza” in Messina. Subsequently he started his training in Pharmaceutical Chemistry and Technology at the Faculty of Pharmacy of the University of Messina, where he received his MSc degree in 2010.

Immediately after the completion of this undergraduate training in 2009, he joined the Division of Pharmacology of the Leiden Academic Centre for Drug Research (LACDR) under the supervision of Prof. Oscar Della Pasqua for the Task Force in Europe for Drug Development for the Young (TEDDY), a network of Excellence funded under the Sixth EU Framework Programme for Research and Technological Development. During this period he contributed to the evaluation of pharmacogenetic factors on the treatment of neuroblastoma as well as on the relevance of modelling and simulation to paediatric drug development.

In 2011 he obtained the license to practice as a pharmacist in Italy and started his PhD research programme at the LACDR under the supervision of Prof. Meindert Danhof and Prof. Oscar Della Pasqua, which led to this thesis. Within his PhD programme he was responsible for the planning, design and implementation of the DEEP-1 multi-centre pharmacokinetic study of deferiprone in children younger than 6 years of age, which was successfully concluded in February 2014.

Since February 2015 he works as a modeller consultant for Quantitative Solutions B.V.

List of Publications

Bellanti F, Del Vecchio GC, Putti MC, Cosmi C, Fotzi I, Bakshi SD, Danhof M, and Della Pasqua O. Model-based optimisation of deferoxamine chelation therapy. *Pharm Res.* [In press]

Bellanti F, van Wijk RC, Danhof M, Della Pasqua O. Integration of PKPD relationships into Benefit Risk Analysis. *Br J Clin Pharmacol.* 2015 May 5 [Epub ahead of print].

Bellanti F, Danhof M, Della Pasqua O. Population pharmacokinetics of deferiprone in healthy subjects. *Br J Clin Pharmacol.* 2014 Dec; 78(6):1397-406.

Bellanti F, Kågedal B, Della Pasqua O. Do pharmacokinetic polymorphisms explain treatment failure in high-risk patients with neuroblastoma? *Eur J Clin Pharmacol.* 2011 May;67 Suppl 1:87-107

Bellanti F, Della Pasqua O. Modelling and simulation as research tools in paediatric drug development. *Eur J Clin Pharmacol.* 2011 May;67 Suppl 1:75-86

