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

Bellanti, F.

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 of 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
Deferoxamine PK model $\frac{dA(1)}{dt} = A(2) \times Q/V2 - A(1) \times Q/V1 - A(1) \times CL/V1$ $\frac{dA(2)}{dt} = A(1) \times Q/V1 - A(2) \times Q/V2$	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
Deferoxamine PKPD model $\frac{dFERRITIN}{dt} = Kin + CRT - Kout \times FERRITIN \times (1 + DFO)$ $CRT = SCL \times e^{-SHP \times FERRITIN}$ $DFO = SLP \times SCss^{AV}$	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 $SCss^{AV}$ is the steady state concentrations
Diabetes and Hypothyroidism hazard model $S(t) = e^{-\int_0^t h(t)dt}$ $h(t) = h_0(t) * e^{\lambda frth + \lambda frtl}$	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
Arthralgia/myalgia logistic model $P = \frac{Css^{\gamma}}{(PC_{50}{}^{\gamma} + Css^{\gamma})}$	Css 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

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.

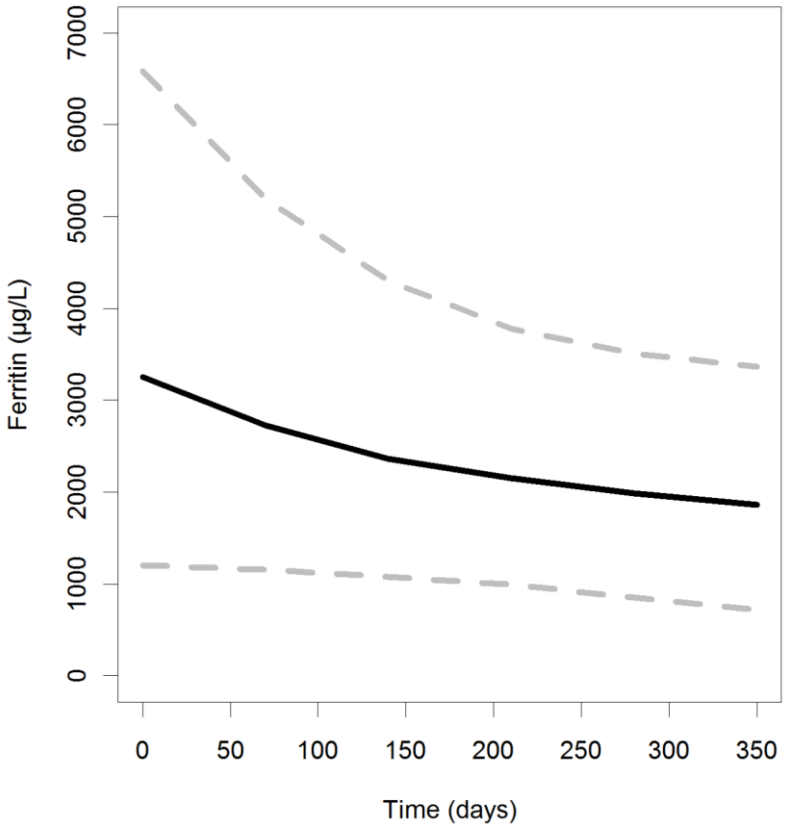

Figure 1. Simulated serum ferritin profiles over a period of 1 year for the Phase III trial in thalassaemic paediatric patients. The solid black line represents the median, whereas the dashed grey lines represent the 5th and 95th percentiles.

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.

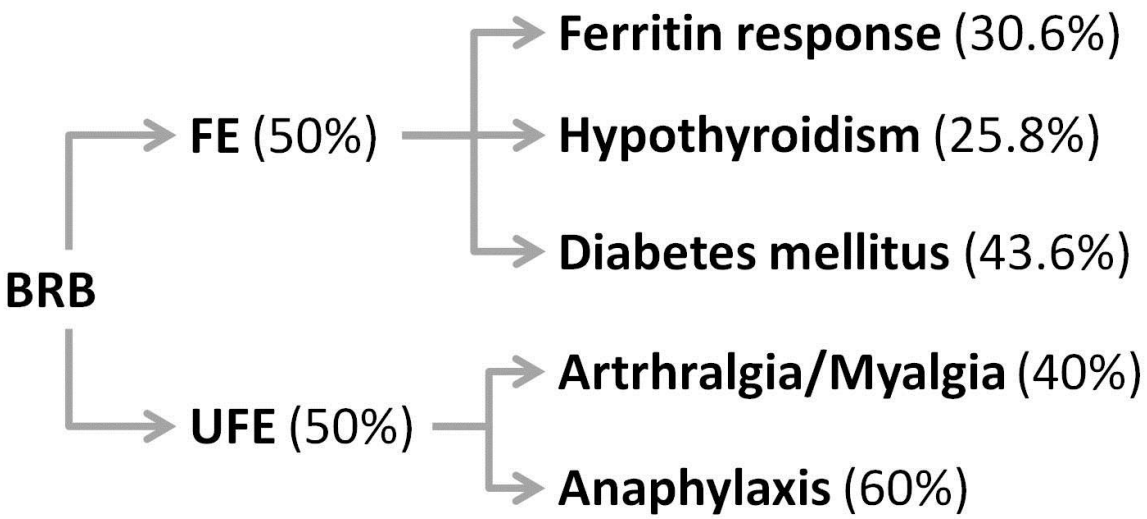

Figure 2. Final value tree and relative weights for the different criteria after discussion with experts. Favourable effects (FE) and unfavourable effects (UFE) were given the same importance whereas among the FE and UFE, diabetes and anaphylaxis were given the major importance respectively.

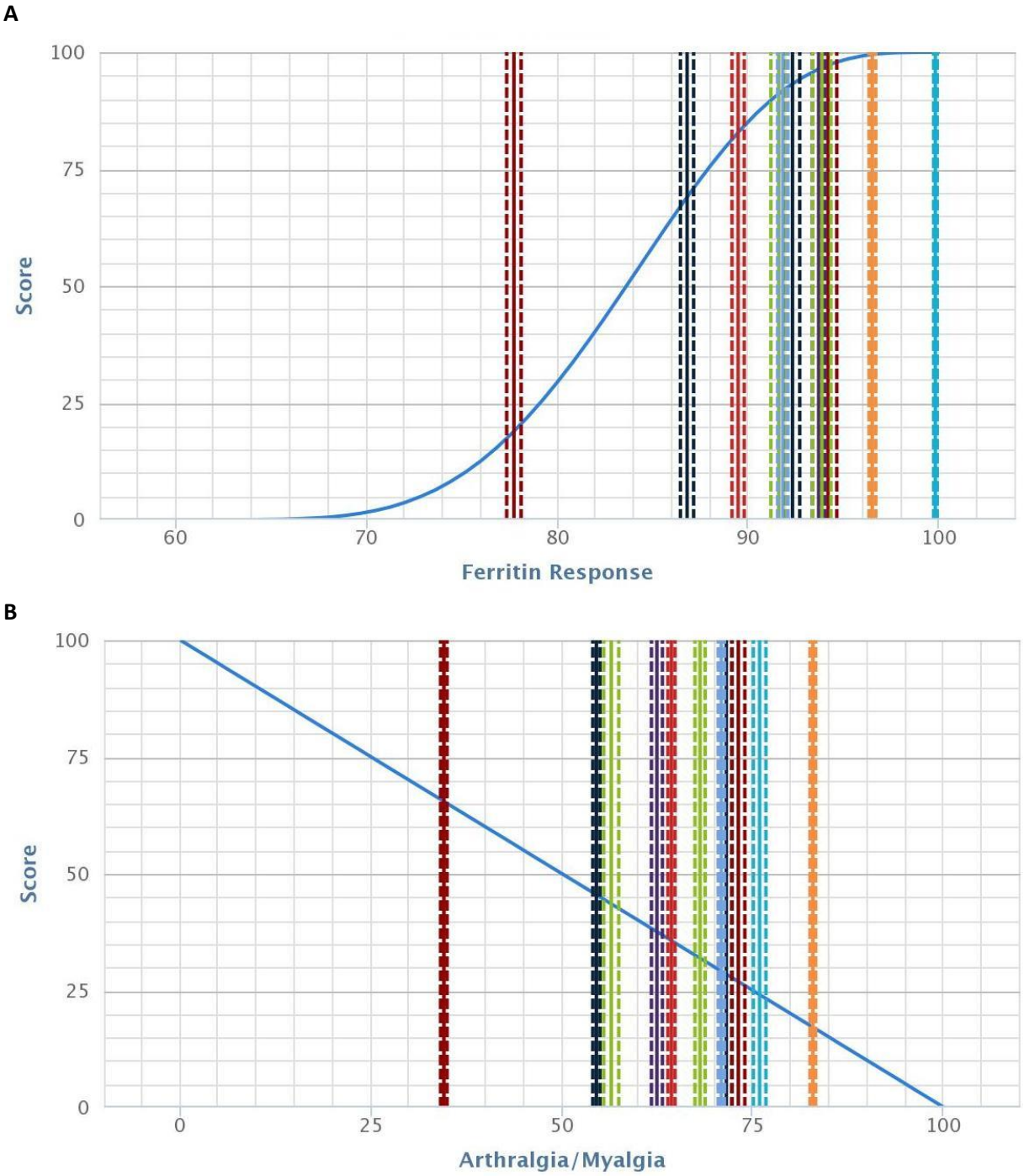

Figure 3. Assessed preference values for two criteria based on the discussion with experts. Panel A shows the non-linear utility function defined for ferritin response, whereas panel B shows the linear utility function defined for arthralgia/myalgia. On the y axis the score is presented in percentage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

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:

$$S_i = w_1 s_{i1} + w_2 s_{i2} + \ldots + w_n s_{in} \qquad \text{Equation 1}$$

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.

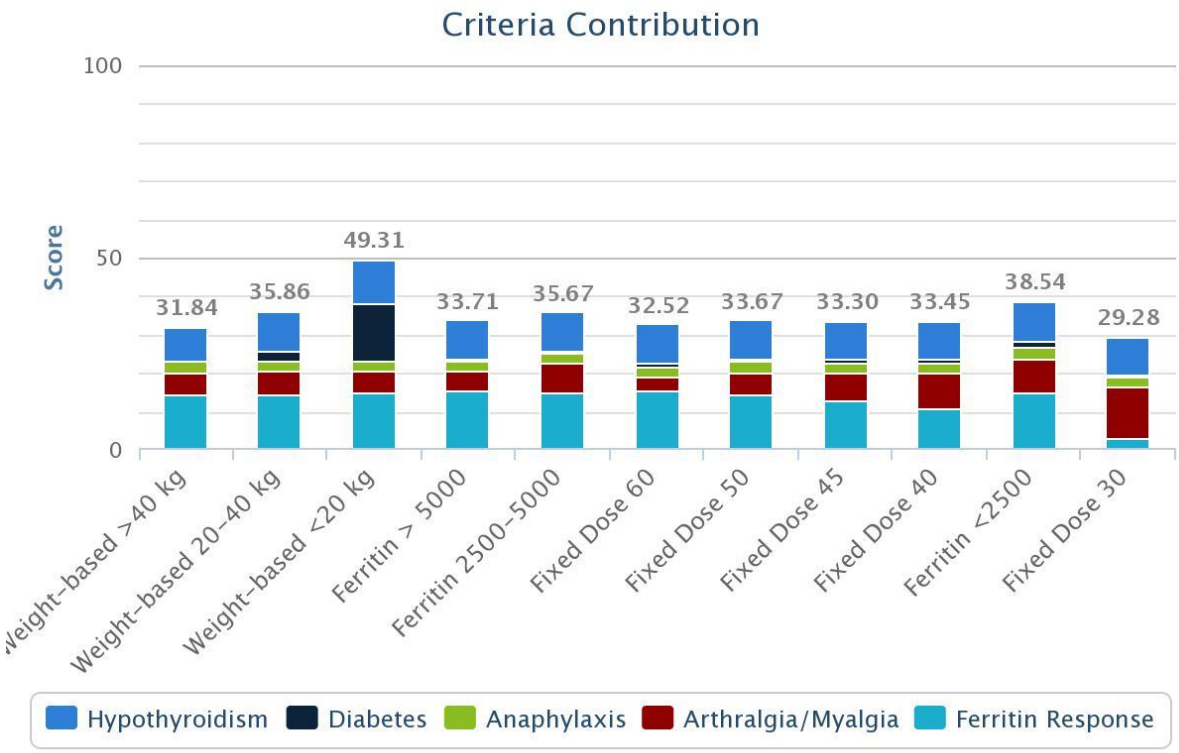

Figure 4. Criteria contribution for the 1 year scenario. The overall weighted score is presented for the different options (the higher the score the better the overall performance of the option appraised). Individual criteria contribution are displayed for each option: light blue, dark red, green, dark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

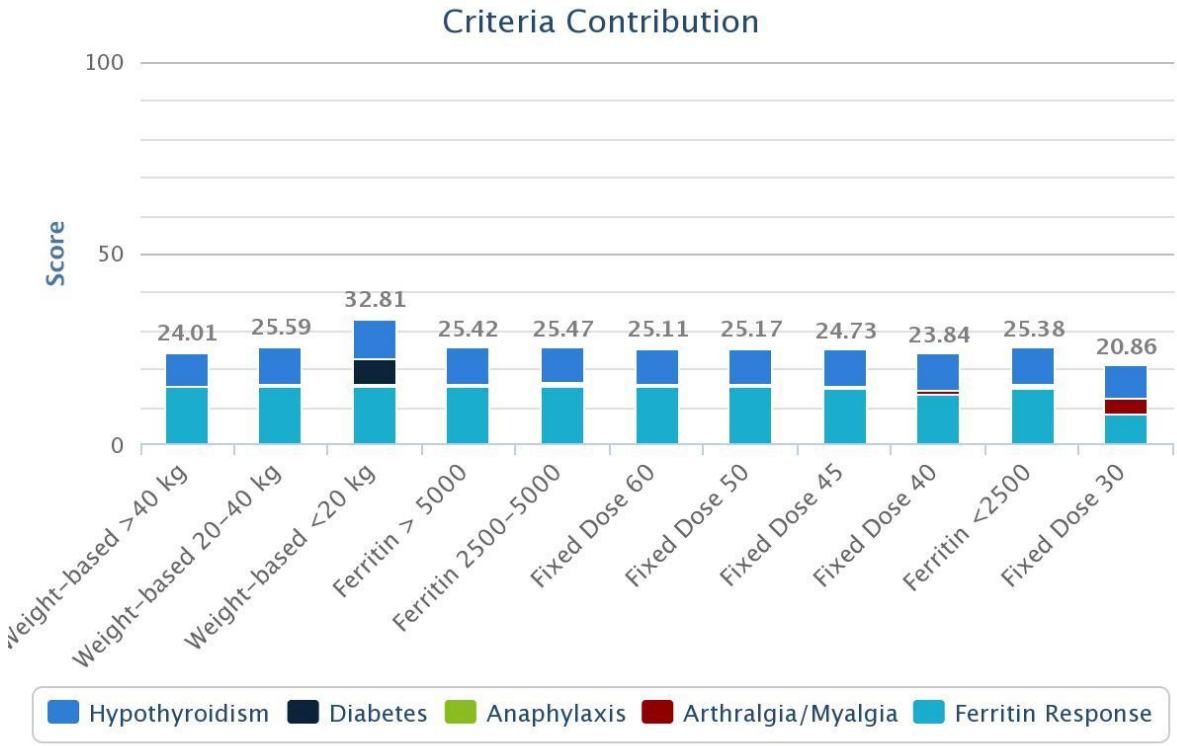

Figure 5. Criteria contribution for the 5 year scenario. The overall weighted score is presented for the different options (the higher the score the better the overall performance of the option appraised). Individual criteria contribution are displayed for each option: light blue, dark red, green, dark 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.

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

	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

	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