
Exposure-response modelling approaches for determining optimal dosing rules in children

Journal Title
XX(X):1–16
©The Author(s) 0000
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/



Ian Wadsworth^{1,2}, Lisa V. Hampson³, Björn Bornkamp³ and Thomas Jaki¹

Abstract

Within paediatric populations there may be distinct age groups characterised by different exposure-response relationships. Several regulatory guidance documents have suggested general age groupings. However, it is not clear whether these categorisations will be suitable for all new medicines and in all disease areas. We consider two model-based approaches to quantify how exposure-response model parameters vary over a continuum of ages: Bayesian penalised B-splines and model-based recursive partitioning. We propose an approach for deriving an optimal dosing rule given an estimate of how exposure-response model parameters vary with age. Methods are initially developed for a linear exposure-response model. We perform a simulation study to systematically evaluate how well the various approaches estimate linear exposure-response model parameters and the accuracy of recommended dosing rules. Simulation scenarios are motivated by an application to epilepsy drug development. Results suggest that both bootstrapped model-based recursive partitioning and Bayesian penalised B-splines can estimate underlying changes in linear exposure-response model parameters as well as (and in many scenarios, better than) a comparator linear model adjusting for a categorical age covariate with levels following ICH E11 groupings. Furthermore, though the Bayesian penalised B-splines approach consistently estimates the intercept and slope more accurately than the bootstrapped model-based recursive partitioning. Finally, approaches are extended to estimate E_{max} exposure-response models and are illustrated with an example motivated by an in vitro study of cyclosporine.

Keywords

Bayesian penalised B-splines, Dosing rules, Exposure-response modelling, Model-based recursive partitioning, Paediatric

1 Introduction

Children of different ages given a new medicine may be characterised by different dose-exposure and exposure-response (E-R) relationships due to age related differences in growth, development and physiological differences¹. Several regulatory guidance documents have suggested general age groupings, such as the International Conference on Harmonisation (ICH) E11 document¹, which suggests one possible categorisation: preterm newborn infants; term newborn infants (0 to 27 days); infants and toddlers (28 days to 23 months); children (2 to 11 years); and adolescents (12 to 16-18 years, depending on region). The National Institute of Child Health and Human Development (NICHD) guideline, suggests similar age groups, but with extra splits at 1 and 6 years. This paper aims to estimate the E-R relationship in children and to identify age groupings which define practical and effective dosing

rules.

An understanding of how the E-R relationship of a drug varies with age will inform whether and how we leverage adult data to support drug development in children. Hampson et al.² reviewed paediatric investigation plans (PIPs) and found that it was common to plan to identify paediatric doses by matching target adult exposures. This is an appropriate dose-finding strategy if E-R relationships are similar

¹Department of Mathematics & Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF, UK.

²Phastar, Macclesfield, UK.

³Advanced Methodology & Data Science, Novartis Pharma AG, Basel, Switzerland.

Corresponding author:

Lisa V. Hampson. Advanced Methodology & Data Science, Novartis Pharma AG, Basel, Switzerland.

Email: lisa.hampson@novartis.com

in adults and children. This assumption might be justified for some paediatric subgroups but not others. For example, Takahashi et al.³ concluded that whilst pubertal (12 to 18 years) and adult patients had similar PD responses to long-term warfarin therapy, there were differences between pre-pubertal (1 to 11 years) patients versus pubertal and adult patients. If E-R relationships can be assumed to be similar across age groups, it may be appropriate to make a complete extrapolation of efficacy data from one age group to another, so that only dose-exposure data are needed in the unstudied age group to identify doses producing exposures known to be efficacious in the studied age group^{2,4}. However, if E-R relationships cannot be considered similar, a partial extrapolation approach⁴ may be considered, where dose-exposure and E-R data may be accrued in specified age groups to confirm differences in E-R relationships and confirm dosing.

One common approach to modelling nonlinear E-R relationships is the Emax model⁵. Thomas et al.⁶ show that the Emax model provides good fit to the dose-response relationship of almost all compounds and diseases in the time window they studied. Parkinson et al.⁷ developed a sigmoid Emax model for the relationship between dapagliflozin exposure and urinary glucose excretion for adult and paediatric patients with type 2 diabetes mellitus. After accounting for significant covariates (e.g. sex, race, baseline fasting plasma glucose), further covariates were included for paediatric patients which failed to improve model fit. The authors took this as evidence that adult and paediatric patients had similar E-R relationships. Earp et al.⁸ used E-R modelling and exposure matching analyses to estimate paediatric doses for esomeprazole for the treatment of gastroesophageal reflux disease. The authors modelled E-R relationships of intragastric pH for adults and children separately and concluded similarity of E-R based on a visual inspection of fitted E-R relationships. In this paper, a more quantitative approach to evaluating differences between E-R relationships is taken using sophisticated modelling approaches.

Age groups characterised by different E-R relationships can be considered as distinct subgroups. Lipkovich et al.⁹ reviewed methods for the identification and analysis of subgroups in clinical trials. Ondra et al.¹⁰ reviewed methods for designing and analysing

clinical trials that aim to investigate differences in treatment effects across subgroups. In this paper, we consider two model-based approaches to quantifying how E-R model parameters vary over a continuous age range: Bayesian penalised B-splines¹¹, and model-based recursive partitioning (MOB)^{12,13} which is used to fit model-based trees to bootstrapped samples of the E-R data. Based on estimates of how E-R model parameters vary with age, we propose an approach to identify the age groups and exposure levels that define a dosing rule which is optimal for targeting a certain level of response; definition of the dosing rule is then completed by using the exposure levels and estimated dose-exposure relationship to make dosing recommendations for each age group. The estimated dose-exposure relationship is not considered in this paper.

Thomas et al.¹⁴ use MOB to estimate patient subgroups with different dose-response curves, and apply this method to data from a dose-finding trial. In this paper, we focus on estimating age groups with different E-R relationships since in practice, when seeking to relate dose to response, a two-step process relating dose to exposure then exposure to response is often adopted. For example, the ICH E4 guidance¹⁵ states that E-R information can help to identify a range of concentrations likely to lead to a satisfactory response, which can in turn inform dose selection. While parameters of the dose-exposure relationship are expected to depend on age, for some medicines parameters of the E-R relationship are expected to remain stable across age groups. In such cases, the two-step modelling process can be advantageous because it enables separate modelling of the dose-exposure and E-R relationships, which allows for changes due to age to be captured in each relationship separately. In a simulation study to compare the performance of the two-step and single stage approaches to dose finding, Berges and Chen¹⁶ found that the two-step approach resulted in more precise E-R model parameter estimation and more accurate dose selection, although gains depend on properties of the drug, trial design features and the target response.

Pharmacokinetics is the study of the time course of drug levels in the body and the mathematical modelling of such data¹⁷. Population-PK models are an extension of PK modelling, studying PK at the

population level and modelling data from all individuals simultaneously¹⁸. Hsu¹⁹ found that in scenarios with increased intrinsic pharmacokinetic (PK) variability, E-R modelling has advantages for dose selection over dose-response modelling, provided measurement error for exposures is small. As an example of a two-stage approach to selecting a dosing rule, Schoemaker et al.²⁰ developed a population PK model to describe the relationship between brivaracetam dose and plasma concentration in adults with partial onset seizures, and a population pharmacokinetic-pharmacodynamic model to describe the relationship between brivaracetam plasma concentration and daily seizure counts. The authors then simulated from these models to estimate the relationship between dose and response, enabling them to identify a dose range producing the maximum response.

This paper proceeds as follows. Section 2 gives a motivating example while Section 3 defines two E-R models. In Section 4, we introduce the methods that will be used to estimate parameters of E-R relationships. Section 5 proposes an approach for using fitted E-R models to identify practical dosing rules for children. We use simulation to evaluate the performance of E-R modelling approaches and the operating characteristics of the dosing rule algorithm. The design of the simulation study is described in Section 6 and the results are presented in Section 7. An example illustrating how the E-R modelling approaches can be applied to non-linear models is given in Section 8. The paper concludes with a discussion in Section 9.

2 Motivating example

We motivate the work that follows by considering the development of epilepsy medicines for paediatric patients with partial onset seizures. Girgis et al.²¹ study both monotherapy and adjunctive therapy with the anti-epileptic drug topiramate, whilst Nedelman et al.²² consider adjunctive therapy with oxcarbazepine. For adjunctive therapy, Girgis et al.²¹ and Nedelman et al.²² take response, $Y = \log\{Z + 110\}$, to be the log-transformed percent change from baseline in seizure frequency, where Z is the percent change from baseline in seizure frequency. The response, Y , is assumed to be normally distributed and a linear function of exposure, measured by the average steady-state trough concentration (C_{min}). Girgis et al.²¹ and Nedelman

et al.²² evaluate the similarity of E-R relationships in adults and children on adjunctive therapy with the aim of justifying the use of extrapolation to support the approval of monotherapy in children. The models and the parameter estimates²¹ will be used to inform the design of realistic simulation scenarios.

3 Exposure-Response models

We start by considering a linear model for the E-R relationship. Suppose E-R data are available from a single study which recruited children aged 0 to 18 years and let Y_i represent the response of subject i , for $i = 1, \dots, N$. If the E-R relationship does not depend upon age, we could model it as:

$$Y_i = \gamma_0 + \sum_{p=1}^P \gamma_p x_{pi} + \gamma_C C_i + \epsilon_i,$$

where C_i is a measure of drug exposure (such as C_{min}), x_{1i}, \dots, x_{Pi} are other covariates influencing response (such as body weight), and $\epsilon_i \sim N(0, \sigma^2)$ is a random error term. We consider the situation where the E-R relationship may differ between age groups, that is, γ_0 and γ_C are functions of age (A):

$$Y_i = \gamma_0(A_i) + \sum_{p=1}^P \gamma_p x_{pi} + \gamma_C(A_i) C_i + \epsilon_i. \quad (1)$$

In Section 4 we will consider different approaches for parameterising $\gamma_0(A_i)$ and $\gamma_C(A_i)$.

Non-linear Emax models are often used to represent the E-R relationship²³. For example, it could be modelled by a sigmoid Emax model:

$$Y_i = \gamma_0(A_i) + \sum_{p=1}^P \gamma_p x_{pi} + \frac{E_{max}(A_i) C_i^{\delta(A_i)}}{EC_{50}(A_i)^{\delta(A_i)} + C_i^{\delta(A_i)}} + \epsilon_i, \quad (2)$$

where for subject i , aged A_i years old, $\gamma_0(A_i)$ is the intercept, $E_{max}(A_i)$ is the maximum effect attributable to the drug, $EC_{50}(A_i)$ is the concentration of the drug that produces half of the maximum effect, and $\delta(A_i)$ (the Hill parameter) governs slope steepness. Here, four of the model parameters may potentially depend upon age.

4 Estimating the exposure-response relationship

In this section we describe three E-R modelling approaches that can be applied when we assume the E-R relationship follows model (1) with age-dependent intercept and slope. These methods are: linear regression with categorical covariates for age groups; MOB and partially additive linear model (PALM) trees; and Bayesian penalised B-splines. We highlight where methods can be applied more generally with non-linear E-R models. A worked example illustrating how each method can be applied to fit a linear E-R model is given in Supplementary Appendix A.

4.1 Linear model fit with categorical age covariates

If we knew that the age groups defined by different E-R relationships were $(a_0 = 0, a_1], (a_1, a_2], \dots, (a_{H-1}, a_H = 18]$, we could define a linear model for the E-R relationship as follows:

$$Y_i = \gamma_0 + \sum_{p=1}^P \gamma_p x_{pi} + \gamma_C C_i + \sum_{h=2}^H \mathbf{I}_{\mathcal{A}_h}(A_i) \{ \gamma_{A,h} + \gamma_{I,h} C_i \} + \epsilon_i, \quad (3)$$

where \mathcal{A}_h is the interval $(a_{h-1}, a_h]$; $\mathbf{I}_{\mathcal{A}_h}(A_i)$ is an indicator function (1 if $A_i \in \mathcal{A}_h$, 0 otherwise); $\gamma_{A,2}, \dots, \gamma_{A,H}$ are the main effects of the age groups; and $\gamma_{I,2}, \dots, \gamma_{I,H}$ are the interactions between age group and exposure. Fitting this model permits estimation of age group specific intercepts and slopes. We include this simple model as a benchmark for comparison with other more complex modelling approaches. Unlike the other methods we consider, this approach requires that age groups be pre-specified rather than estimating them from the data.

4.2 MOB and PALM trees

Building on model (3), MOB allows data to be split into groups based on partitioning variables, with each subgroup characterised by its own parametric model¹³. We implement MOB using age as the only partitioning variable. The MOB algorithm we use comprises the following steps¹³: Fit a parametric model to the dataset, finding parameter estimates by minimising an objective function; test whether the model parameters

significantly change with age using a generalized M-fluctuation test^{13,24} which assesses whether the scores of the model systematically deviate from 0 with age; partition the model into two subgroups with respect to age by finding the value of age which minimises an objective function segmented at this split point; repeat the fitting, testing and splitting procedure in each identified age group until no significant changes are found in the model parameters over age within each group. In our subsequent examples, the parametric model will be taken to be a linear model, where the parameters of interest are the intercept and slope. The MOB algorithm¹³ can be implemented using the ‘mob’ function found in the ‘partykit’ package^{13,25} in R²⁶. As MOB allows subgroups defined by any parametric model, non-linear models (such as Emax models) are possible.

PALM trees are a variation of MOB, allowing for global parameters which remain constant across subgroups. However, PALM trees are restricted to generalised linear models (GLM)¹². For our linear model example with outcome Y_i and partitioning age variable A_i , PALM trees can contain globally fixed linear effects $\gamma_1, \dots, \gamma_P$ for covariates x_{1i}, \dots, x_{Pi} and subgroup-wise varying linear effects $\gamma_0(A_i)$ and $\gamma_C(A_i)$, as in equation (1). PALM trees use the MOB algorithm described above to identify age groups with distinct GLMs. In order to allow for global parameters which remain constant across age groups, an EM-type algorithm is used. This iterates between estimating the global effects, $\gamma_1, \dots, \gamma_P$, for the current PALM tree and estimating the PALM tree (using the above MOB algorithm) for a given set of global effect estimates, $\hat{\gamma}_1, \dots, \hat{\gamma}_P$. The algorithm can be implemented in R²⁶ using the ‘palmtree’ function found in the ‘partykit’ package^{12,25}. We implement PALM trees with the default tuning parameters, i.e. a significance level of 0.05 and no maximum tree depth. An advantage of tree based methods is the easy to understand output: each final partitioned subgroup of the tree represents an age group, with model parameter estimates given for each group.

We implement MOB and PALM tree approaches using bootstrap aggregating²⁷ to improve the accuracy and precision of age-specific E-R model parameter estimates and reduce overfitting. The E-R data are bootstrapped and each bootstrap sample is used to fit

a MOB or PALM tree. From each bootstrap tree fit, estimates of age-specific model parameters (intercept and slope) can be evaluated for a grid of ages covering the interval $[0, 18]$ years. For each grid point in turn, we then aggregate across the bootstrap samples and, applying linear interpolation to the average age-specific parameter estimates, can thus obtain an estimate of the E-R intercept or slope for any given age. The important aspect to note here is that no parametric assumptions are made about the form of the relationship between each model parameter and age. One disadvantage of this is that these relationships cannot then be easily recorded in a closed form for future reference.

We fit linear E-R models using PALM trees in Section 6 because we also consider the case of having an additional global covariate whose effect is independent of age, which we present in Supplementary Appendix B. In Section 8, we fit non-linear E-R models using MOB.

4.3 Bayesian penalised B-splines

Splines define flexible regression models by joining smooth curves (differentiable at every point) together at knot points²⁸. An E-R model parameter that can be written as a smooth function of A , $f(A)$, can be modelled as a spline. Here, we will consider the penalised B-splines developed by Eilers and Marx¹¹. B-splines can be written as a linear combination of B-spline basis functions of degree d , that is, $B_1(A; d), \dots, B_J(A; d)$:

$$f(A) = \sum_{j=1}^J \beta_j B_j(A; d). \quad (4)$$

A B-spline basis function of degree d consists of $d + 1$ polynomial curves of degree d , each joined in sequence¹¹. The degree of the B-spline basis controls how differentiable the spline is and can influence the smoothness of the spline. We implement B-splines of degree 2 as in the examples we have considered we gain little in terms of smoothness for the added complexity of using degree 3 B-splines. We therefore fit linear E-R models defining the intercept and slope as B-splines of degree 2:

$$\begin{aligned} \gamma_0(A_i) &= \sum_{j=1}^J \beta_{0j} B_j(A_i; d = 2), \\ \gamma_C(A_i) &= \sum_{j=1}^J \beta_{Cj} B_j(A_i; d = 2). \end{aligned}$$

We set $J = 26$ given our choice of degree and number of knots: five equally spaced knots within each of the four ICH E11 age groups (not including pre-term newborn infants), knots at each age group boundary, along with two external knots below age zero and two above 18 years. We use the function ‘splineDesign’ in the R package ‘splines’²⁶ to construct our 26 B-spline basis functions. Further details of how the B-spline basis functions are constructed can be found in Bowman and Evers²⁸. Note that for penalised B-splines, Eilers and Marx¹¹ recommend using equidistant knots and suggest that there are no gains to be made from using unequally spaced knots, as the penalty smooths any sparse areas. However, we specify knots using the prior information on potential age groupings that is contained in the ICH E11 guidance document¹. By specifying an equal number of knot points across each ICH E11 age group, knots are more densely spread across age ranges where model parameters are expected to change most rapidly with age. A sensitivity analysis to explore the impact of knot placement would be appropriate in many cases.

For penalised B-splines, a roughness penalty is used to control the smoothness of the estimated spline, rather than the choice of knot location and number¹¹. In a Bayesian context, penalised B-splines are implemented placing random walk priors on the B-spline coefficients^{28,29}. For example, to penalise differences between adjacent B-spline coefficients, first-order random walk priors are used:

$$\begin{aligned} \beta_{0,j} | \beta_{0,j-1} &\sim N(\beta_{0,j-1}, \tau_0^2), & \text{for } j = 2, \dots, J \\ \beta_{C,j} | \beta_{C,j-1} &\sim N(\beta_{C,j-1}, \tau_C^2), \end{aligned}$$

with $\beta_{0,1} \sim N(0, 100)$ and $\beta_{C,1} \sim N(0, 100)$. This penalises B-spline coefficients by shrinking towards a common constant²⁸, which is desirable in our context since we anticipate that there may be age ranges on which a model parameter is fairly stable followed by periods of rapid change. We stipulate diffuse Inverse-Gamma(1, 0.005) priors for τ_0 and τ_C , similar to Lang and Brezger (2004)²⁹ who place an Inverse-Gamma prior on the variance of the random walk prior. We do not weight τ_0^2 and τ_C^2 by the distance between successive knot points, as suggested by Kneib et al.³⁰, to allow larger prior variation when there are larger steps between knots. This is because in our setting, we have purposefully placed knots closer together over age intervals where the most rapid changes with age are

anticipated.

We fit the Bayesian penalised B-splines model using Hamiltonian Monte Carlo, calling Stan³¹ from R²⁶ using the RStan package³², and running three chains with a default thinning rate of one for 3000 iterations, 1500 of which are discarded as burn-in samples. Following equation (4), the posterior means of the B-spline coefficients are multiplied by the B-spline basis functions to estimate the B-spline for the respective E-R model parameter.

Bayesian penalised B-splines are a very flexible modelling approach, with the capacity to be used to represent the parameters of any parametric E-R model. The ability to write the relationship between E-R parameters and age in a simple form, as in equation (4), means it is easy to record and communicate the estimated relationship. However, Bayesian penalised B-spline models can comprise many parameters which can make them computationally expensive to fit.

5 Dosing recommendations

5.1 Optimisation criterion

We could use the modelling approaches described in Section 4 to derive personalised dosing recommendations tailored to a patient's exact age and baseline covariates. However, for practical reasons, we seek to identify dosing rules based on wider age subgroups. As outlined in Section 1, we focus on identifying age groups and exposure levels targeting a certain level of response, assuming that in a second step we could use a PK model to link each target exposure to dose. Therefore, we use 'dosing rule' as a short-hand to refer to a set of age groupings and corresponding target exposures. First, we derive target exposure levels for up to K age groups of children. For practical reasons, K would likely be small, e.g. $K = 5$ in the ICH E11 guideline¹. When defining the target exposure for each age group, we would like to minimise the difference between the expected response and a target response denoted by Y^* . For the epilepsy example, a 50% change in seizure frequency from baseline would be an appropriate target response, so that $Y^* = \log(-50 + 110)$.

We derive dosing rules assuming the E-R model and parameter estimates (maximum likelihood for the

frequentist approaches, posterior means for the Bayesian penalised B-splines) are identical to the true model and parameter values. Given a proposed age grouping, let C_k denote the target exposure for the k th age group (a_{k-1}, a_k] needed for a patient aged $(a_{k-1} + a_k)/2$ years to have expected PD response equal to Y^* . Furthermore, define $D_a = |\mathbb{E}[Y | A = a, C = C_k] - Y^*|$. If the E-R model adjusts for a set of baseline covariates, expectations of Y are calculated conditioning on average covariate values at age $A = a$, while C_k is calculated for a patient with average covariate values at age $(a_{k-1} + a_k)/2$. One approach would be to find the dosing rule minimising the objective function $F = \int_0^{18} D_a da$, where rules minimising F minimise the total absolute difference between the expected response and Y^* . F weights equally the performance of the dosing rule at every age. This is undesirable in our context since if E-R model parameters do depend on age, it may be reasonable to expect parameters to change rapidly over short intervals (i.e. between 0 - 2 years) and remain fairly stable across the adolescent age range. Minimising F would favour rules which dose most ages effectively, where inaccurate dosing over narrow age intervals would not be seriously penalised. However, our aim is to ensure all ages are dosed appropriately. With this in mind, we choose dosing rules to minimise:

$$G = \frac{1}{a_1^*} \int_0^{a_1^*} D_a da + \frac{1}{a_2^* - a_1^*} \int_{a_1^*}^{a_2^*} D_a da + \dots + \frac{1}{a_P^* - a_{P-1}^*} \int_{a_{P-1}^*}^{a_P^*} D_a da, \quad (5)$$

where $a_1^* < a_2^* < \dots < a_P^*$ are fixed and pre-specified age boundaries and may be based on regulatory guidance, such as the ICH E11 guideline¹ or the NICHD guideline³³. We define these boundaries in line with the NICHD guidelines. Finding dosing rules which minimise G means that we give equal weight to the performance of the dosing rule in a number of paediatric age groups considered as our best prior guesses.

5.2 Identifying an optimal number of age groups in our dosing rule

Define $\Lambda_K = (a_0, \dots, a_K)$ as the vector of age boundaries defining the optimal dosing rule with K groups; \mathcal{C}_K as the vector of target exposures; and G_K^* as the minima of G for K age groups. Furthermore, let K_{max} denote the maximum number of age groups considered to be plausible or workable in practice,

which would be pre-specified based on feedback from clinicians. We use the following algorithm to define a paediatric dosing rule:

1. Begin with $K = 1$ age group;
2. For K age groups, search over configurations of Λ_K to find the dosing rule minimising G_K ;
3. Save G_K^* , Λ_K^* , and \mathcal{C}_K^* ;
4. Repeat steps (2) and (3), successively increasing K by one until $K = K_{max}$.

The minima $G_1^*, \dots, G_{K_{max}}^*$ can be compared to see if increasing K always produces a worthwhile increase in the accuracy of the dosing rule. The optimum value of K , balancing the trade-off between complexity and accuracy, is denoted by K^* . In some scenarios, a more automated approach to selecting K^* is possible. In these cases, for each $K = 1, \dots, 5$, we propose calculating the percentage difference between G_{K+1}^* and G_K^* . The value of K where the percentage change is less than $c = 25\%$ is taken as K^* . The arbitrary choice of c used here is intended to illustrate one possible approach and will be adopted in the simulation study described in the next section.

6 Design of the simulation study

We performed a simulation study to explore the performance of the modelling approaches described in Section 4 and the approach of Section 5 for defining dosing rules. We consider a range of data generation scenarios for the linear model described in Section 3. For the categorical age covariates model, we follow the ICH E11 age groups to fix the age intervals as $\mathcal{A}_1 = (0, 28/365]$, $\mathcal{A}_2 = (28/365, 2]$, $\mathcal{A}_3 = (2, 12]$ and $\mathcal{A}_4 = (12, 18]$ in equation (3), across all scenarios.

We simulate studies enrolling 25 subjects into each of four ICH E11 age groups, $(0, 28/365]$, $(28/365, 2]$, $(2, 12]$, $(12, 18]$, excluding preterm newborn infants. Within age group $(a_{i-1}, a_i]$, the age of patient i is sampled from a $\text{Uniform}(a_{i-1}, a_i)$ distribution. We consider 11 scenarios, as illustrated in Figures 1 and 2, for how E-R model parameters vary with age. More detail on these scenarios is provided in Supplementary Table S1, Supplementary Figure S9 and Appendix A. We only consider scenarios where the E-R intercept and slope change monotonically with age, since these differences are most realistic in the context of the epilepsy example.

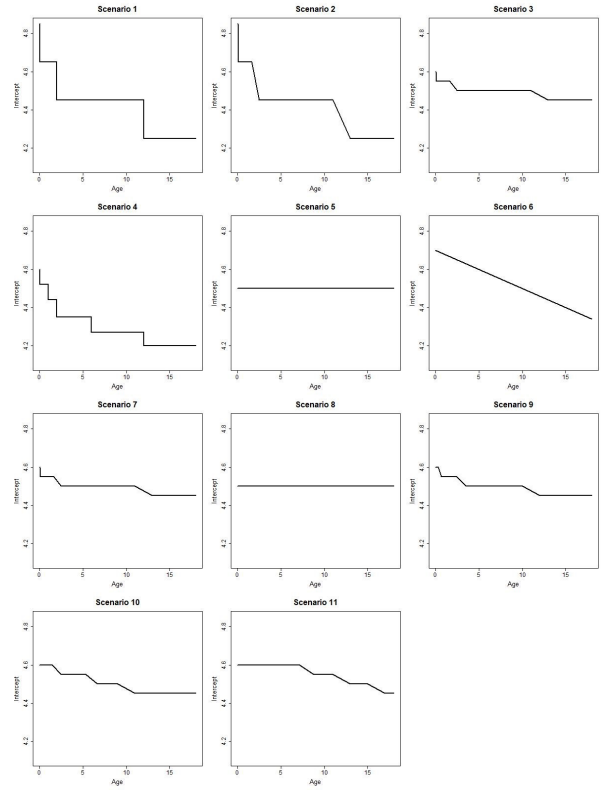


Figure 1. Plot showing how the intercept of the E-R model changes with age in simulation scenarios 1-11.

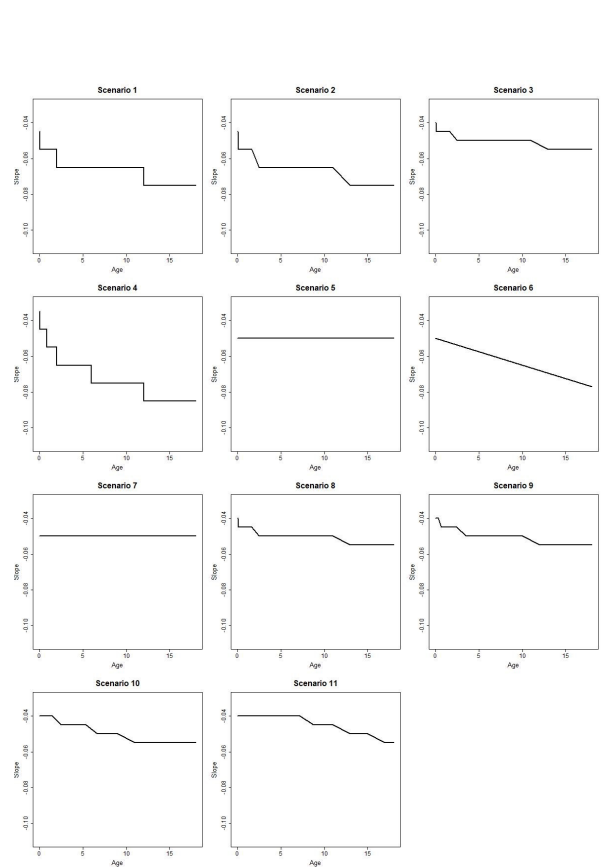


Figure 2. Plot showing how the slope of E-R model changes with age in simulation scenarios 1-11.

We measure exposure by C_{min} . Following Wadsworth et al.³⁴, we sample $\log(C_{min})$ from a $N(\log(2.94), 0.921)$ distribution, truncating samples above by $\log(17.27)$ to avoid excessively high concentrations. We sample random errors from a $N(0, 0.02)$ distribution. These simulated values are used to generate patient responses, Y_i , according to equation (1). We simulate 1000 data sets for each scenario and approach using the statistical software R²⁶.

6.1 Evaluating different approaches to modelling the E-R relationship

We use the following measures to compare the modelling approaches. Define \mathcal{A} as a grid of $Q = 40000$ equally spaced ages between 0 and 18 years. For each age, $A_q \in \mathcal{A}$, we first measure how well each of the methods has estimated the true intercept and slope parameters. We do this by comparing the true parameters, $\gamma_0(A_q)$ and $\gamma_C(A_q)$, with our estimates of the parameters, $\hat{\gamma}_0^{(m)}(A_q)$ and $\hat{\gamma}_C^{(m)}(A_q)$, based on simulated dataset m , for $m = 1, \dots, 1000$. For simplicity, henceforth we will refer to a general E-R model parameter $\gamma^{(m)}(A_q)$ and corresponding estimate $\hat{\gamma}^{(m)}(A_q)$.

Let $\mathbb{E}[\widehat{\hat{\gamma}}(A_q)] = \frac{1}{M} \sum_{m=1}^M \hat{\gamma}^{(m)}(A_q)$. We compute the average absolute bias (AAB), Empirical Standard Deviation (ESD) and Empirical Mean Squared Error (EMSE) of a parameter estimator at age A_q as:

$$\begin{aligned} \text{AAB}(\hat{\gamma}(A_q)) &= \frac{1}{M} \sum_{m=1}^M \left| \hat{\gamma}^{(m)}(A_q) - \gamma(A_q) \right|, \\ \text{ESD}(\hat{\gamma}(A_q)) &= \sqrt{\frac{1}{M-1} \sum_{m=1}^M \left(\hat{\gamma}^{(m)}(A_q) - \mathbb{E}[\widehat{\hat{\gamma}}(A_q)] \right)^2}, \\ \text{EMSE}(\hat{\gamma}(A_q)) &= \frac{1}{M} \sum_{m=1}^M \left(\hat{\gamma}^{(m)}(A_q) - \gamma(A_q) \right)^2, \end{aligned}$$

for $q = 1, \dots, Q$. Using the grids of AAB, ESD and EMSE values thus produced, we use Simpson's rule^{35,36} to calculate the integrated absolute bias, integrated empirical SD and integrated empirical MSE for the E-R parameter estimator. These metrics can be interpreted as overall measures of accuracy, precision and MSE of an estimate of the functional relationship between an E-R model parameter and age.

Similarly, let Y_{qj} denote the response at age, A_q , and exposure, $C_j \in \mathcal{C}$, where \mathcal{C} is a grid of $J = 40000$ equally spaced exposures between 0 and 18. We

wish to compare the estimated expected response at exposure level C_j , $\hat{\mathbb{E}}^{(m)}[Y_{qj}] = \hat{\gamma}_0^{(m)}(A_q) + \hat{\gamma}_C^{(m)}(A_q)C_j$, with the true expected response at C_j given by $\mathbb{E}[Y_{qj}] = \gamma_0(A_q) + \gamma_C(A_q)C_j$.

Let $\widehat{\mathbb{E}}[Y_{qj}] = \frac{1}{M} \sum_{m=1}^M \hat{\mathbb{E}}^{(m)}[Y_{qj}]$. For each $j = 1, \dots, J$, and $q = 1, \dots, Q$ calculate:

$$\begin{aligned} \text{AAB}(\hat{\mathbb{E}}[Y_{qj}]) &= \frac{1}{M} \sum_{m=1}^M \left| \hat{\mathbb{E}}^{(m)}[Y_{qj}] - \mathbb{E}[Y_{qj}] \right|, \\ \text{ESD}(\hat{\mathbb{E}}[Y_{qj}]) &= \sqrt{\frac{1}{M-1} \sum_{m=1}^M \left(\hat{\mathbb{E}}^{(m)}[Y_{qj}] - \widehat{\mathbb{E}}[Y_{qj}] \right)^2}, \\ \text{EMSE}(\hat{\mathbb{E}}[Y_{qj}]) &= \frac{1}{M} \sum_{m=1}^M \left(\hat{\mathbb{E}}^{(m)}[Y_{qj}] - \mathbb{E}[Y_{qj}] \right)^2. \end{aligned}$$

These evaluations produce $Q \times J$ matrices of values for AAB, ESD, EMSE. For each C_j , for $j = 1, \dots, J$, we then numerically integrate over age using Simpson's rule, and then apply Simpson's rule again to integrate over exposure to obtain the integrated absolute bias, integrated empirical SD and integrated empirical MSE for a patient's expected response. These can be interpreted as overall measures of the accuracy, precision and MSE of our estimate of the E-R relationship across a continuum of ages.

6.2 Measuring the accuracy of dosing rules

Following the algorithm of Section 5, we find dosing rules comprising $K = 1, \dots, 6$ age groups, with associated target exposures and minimum objective function values. We want to assess the performance of this dosing rule identification process. For the m^{th} simulated dataset we first take the derived K 'optimal' age groups, $(a_0^{(m)} = 0, a_1^{(m)}], \dots, (a_{K-1}^{(m)}, a_K^{(m)} = 18]$, and estimates of corresponding target exposure levels, $\hat{C}_1^{(m)}, \dots, \hat{C}_K^{(m)}$, and evaluate the true expected response, at the target exposure levels, according to the simulation model. That is, at age $A_q \in \mathcal{A}$, we define

$$\hat{\mathbb{E}}^{(m)}[Y_{qK}] = \sum_{k=1}^K \mathbf{I}_{\mathcal{A}_k^{(m)}}(A_q) \left[\gamma_0(A_q) + \gamma_C(A_q) \hat{C}_k^{(m)} \right],$$

for $q = 1, \dots, Q$, where $\mathcal{A}_k^{(m)}$ is the interval $(a_{k-1}^{(m)}, a_k^{(m)})$ and $\mathbf{I}_{\mathcal{A}_k^{(m)}}(A_q)$ is an indicator function, which takes the value 1 if $A_q \in \mathcal{A}_k^{(m)}$ and 0 otherwise. This measure is the true expected response, under the simulation model, implied by the estimated dosing rule. Comparing this to the target response will allow us to measure the

accuracy of our dosing rule. For each $q = 1, \dots, Q$ and $K = 1, \dots, K_{max}$ we find $Y_{qK,diff}$, the absolute difference between $\hat{\mathbb{E}}^{(m)}[Y_{qK}]$ and Y^* averaged over the 1000 simulated datasets:

$$Y_{qK,diff} = \frac{1}{M} \sum_{m=1}^M \left| \hat{\mathbb{E}}[Y_{qK}^{(m)}] - Y^* \right|.$$

This measure can be interpreted as the accuracy of the K -group optimal dosing rule at age A_q . As with Section 6.1, we calculate the integral of $Y_{qK,diff}$ over age using Simpson's integration. This measure gives an overall measure of the accuracy of the K -group optimal dosing rule and allows us to evaluate how close the true expected response (derived from the simulation model) is to the target response when children are dosed according to the estimated optimal dosing rule. We also consider how many of the simulated datasets would lead us to select a dosing rule with $K^* = 1, \dots, K_{max}$ groups, in order to evaluate the typical complexity of optimal dosing rules and how this varies with the extent of differences between E-R model parameters across age groups.

7 Results

Figures 3–5 plot the integrated absolute bias and integrated empirical SD of E-R model parameter estimators for each modelling approach in each simulation scenario. For estimates obtained fitting Bayesian penalised B-splines, bootstrapped PALM trees, a single PALM tree and the linear model with categorical age covariate, Supplementary Tables S2–S5, in Supplementary Appendix C, present the integrated average absolute bias, empirical SD (as shown in Figures 3–5) and empirical MSE (not included in the paper) of the estimated intercepts, slopes and expected response.

Comparing different modelling approaches within a scenario, Figures 3–5 suggest that, in general, estimates of the functional relationship between the E-R model intercept and slope parameters obtained via Bayesian penalised B-splines are more accurate than estimates obtained using bootstrapped PALM trees. The single PALM tree fit is outperformed by the bootstrapped PALM tree approach in terms of both integrated absolute bias and empirical SD across most scenarios and both parameters, suggesting that bootstrapping is a refinement to the single PALM tree approach. As would

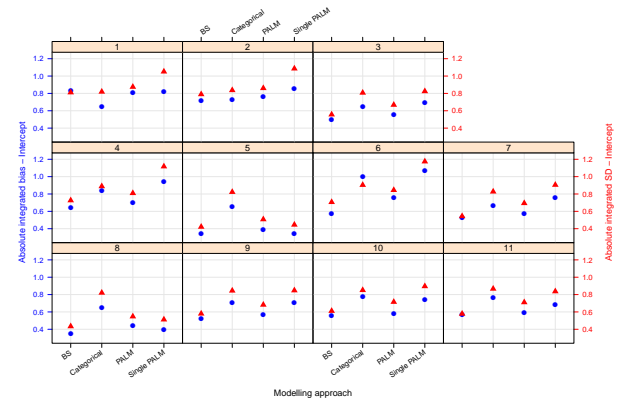


Figure 3. Integrated absolute bias (blue circles) and integrated empirical SD (red triangles) for the E-R model intercept. On the horizontal axis, 'BS' refers to the Bayesian penalised B-splines approach, 'Categorical' the linear model adjusted for a categorical age covariate, and 'PALM' and 'singlePALM' label the bootstrapped PALM tree approach and single PALM tree, respectively.

be expected, the categorical covariate fit performs best in terms of accuracy and precision in scenario 1, where age groups are most distinct and follow the categories suggested by the ICH E11 guidance, excluding pre-term newborns.

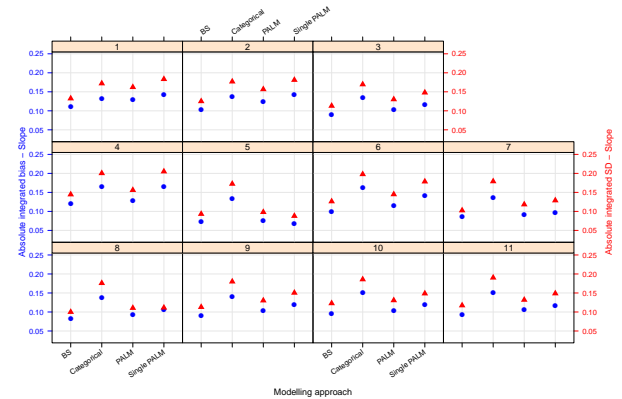


Figure 4. Integrated absolute bias (blue circles) and integrated empirical SD (red triangles) for the slope of the E-R model. On the horizontal axis, 'BS' refers to the Bayesian penalised B-splines approach, 'Categorical' the linear model adjusting for a categorical age covariate, and 'PALM' and 'singlePALM' label the bootstrapped PALM tree approach and single PALM tree, respectively.

Figure 6 compares the performance of dosing rules minimising G_K under different values of K , derived from E-R models fitted using different modelling approaches. As the linear model adjusting for a categorical age covariate has fixed age groups and the single PALM tree approach estimates specific age groupings, results for optimised dosing rules are only

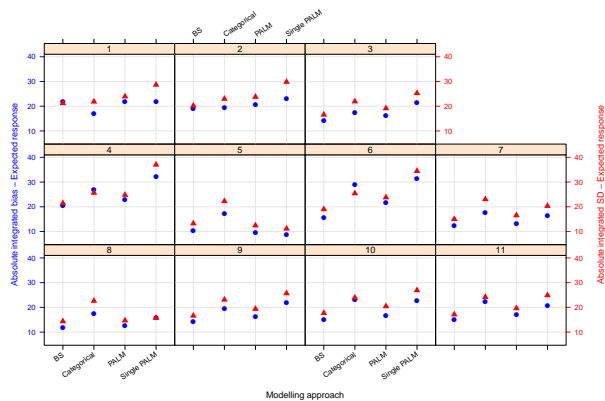


Figure 5. Integrated absolute bias (blue circles) and integrated empirical SD (red triangles) for the expected response. On the horizontal axis, 'BS' refers to the Bayesian penalised B-splines approach, 'Categorical' the linear model adjusting for a categorical age covariate, and 'PALM' and 'singlePALM' label the bootstrapped PALM tree approach and single PALM tree, respectively.

presented for the Bayesian penalised B-splines and bootstrapped PALM tree approaches. Figure 6 shows that overall both Bayesian penalised B-splines and bootstrapped PALM trees define K -group dosing rules with a similar performance in terms of getting the expected response close to the target response under the simulation model. In most scenarios, there comes a point at which there is little to be gained in terms of accuracy by refining the dosing rule further by allowing for additional age groups. As K increases, typically either the true expected response (under the simulation model and implied by the estimated dosing rule) better matches the target response or differences between the performances of the K -group dosing rules for other modelling approach diminish.

Figure 7 shows the percentage of simulations where the global optimum dosing rule comprises K^* age groups for various values of K^* . It is important to note, however, that the dosing rule age groups determined using the algorithm defined in Section 5.2 may not necessarily be identical to the true underlying E-R age groups, as if there are large differences in expected response between underlying E-R age groups, a better fit may be achieved by dosing rules splitting age groups around big changes and combining age groups with smaller changes. However, it is interesting to explore the values of K^* defining the global optimal dosing rules to assess their complexity. Additionally, the complexity of the derived dosing rules will depend on the quantitative threshold used to identify K^* described in Section 6.2;

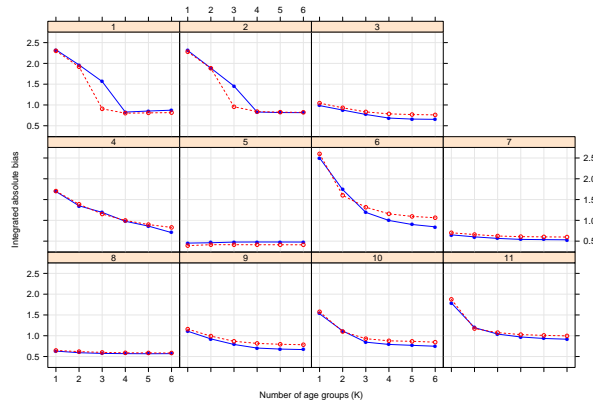


Figure 6. Integrated absolute difference between the target response and true expected response when children are dosed according to the K group optimal dosing rule. Results are shown for dosing rules obtained modelling the E-R relationship using Bayesian penalised B-splines (solid blue line) and bootstrapped PALM trees (dashed red line).

with a different threshold, c , dosing rules with different K^* may be selected as optimal. We find optimal dosing rules minimising G to be cautious, forming slightly more age groups than the underlying E-R age groups.

Focusing on Bayesian penalised B-splines, we see from Figure 7 that in scenario 1, where larger differences are present in the underlying E-R model parameters, the large majority (81.6%) of simulated datasets would lead to the investigator selecting a global optimum dosing rule with $K^* = 4$, as would a smaller majority (54%) of datasets in scenario 3. This suggests that when underlying E-R relationships across age groups become less distinct, dosing rules with smaller K^* are selected. In scenario 4, the majority of simulated datasets would lead to the investigator selecting global optimum dosing rules with $K^* = 4$, although there is a trend to larger K^* compared with other scenarios. In scenario 5, where underlying E-R model parameters do not depend on age, a higher percentage of datasets lead to the selection of a dosing rule defined by a smaller K^* .

Similar patterns are seen for the bootstrapped PALM trees approach in Figure 7. It seems that both bootstrapped PALM trees and the Bayesian penalised B-splines approach are capable of identification of dosing rules with multiple age groups when differences in the underlying E-R relationships across age groups are large, but fewer are identified as differences diminish. For the single PALM tree fit, for scenarios where larger differences are present in the underlying E-R model

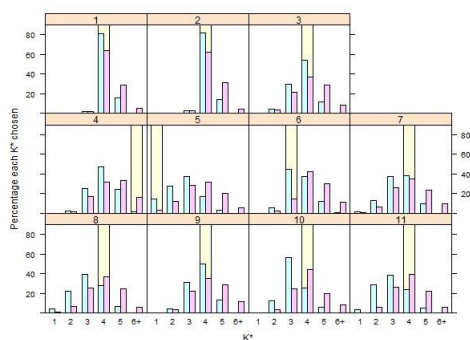


Figure 7. Percentage of 1000 simulations in which K^* , the optimal number of age groups in the dosing rule, takes each value shown. K^* is selected according to the algorithm described in Section 5.2 for Bayesian penalised B-spline (blue) and bootstrapped PALM tree (pink) approaches. The values of K^* chosen by applying the algorithm in Section 5.2 to the true underlying E-R relationships in each scenario are shown by the yellow bars.

parameters, as in scenarios 1 and 2, a single PALM tree often identifies dosing rules with four groups; 96.9% and 94.6% would choose four groups, respectively. In not one scenario did a single PALM tree select a dosing rule with $K^* > 4$.

8 Extension to Emax model

We consider a simulated example informed by the data presented in Marshall and Kearns³⁷, who modelled the relationship between cyclosporine concentration and in vitro inhibitory effect on peripheral blood monocyte (PBM) proliferation as a sigmoid Emax curve (2). We simulate responses for 41 subjects assigned to one of four age groups: 10 infants (0–1 year); 12 children (1–4 years); 9 pre-adolescents (4–12 years); and 10 adults (12–18 years). Data are generated such that for each of the following concentrations of cyclosporine (6.25; 12.5; 25; 50; 100; 250; 500; 1000; and 5000 ng/mL) a patient was recruited from each age group and the remaining patients in each age group (1 infant; 3 children; 1 adult) were randomly assigned a concentration from this set. Within an age group, patients' ages are assumed to follow a uniform distribution. In a deviation from Marshall and Kearns³⁷, patient responses are simulated according to a hyperbolic Emax model (setting $\delta(A) \equiv 1$), although we follow the original publication to force a zero intercept ($\gamma_0(A) \equiv 0$). Patient responses are simulated setting the remaining EC50 and Emax model parameters equal to the age group specific parameter

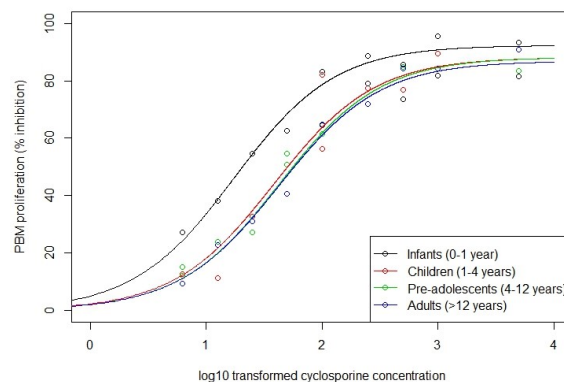
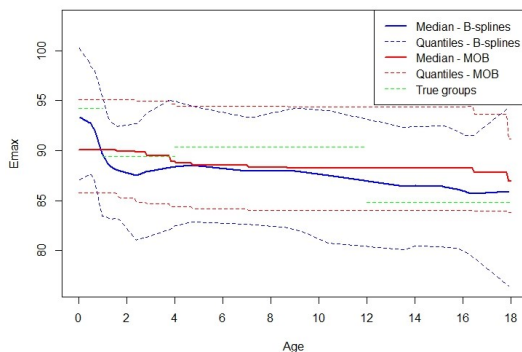


Figure 8. Fitted curves of the relationship between log base-10 transformed cyclosporine concentrations and PBM proliferation based on frequentist two parameter Emax model fit for each of the four age groups considered. Fitted curves are the solid lines and the points are simulated data.

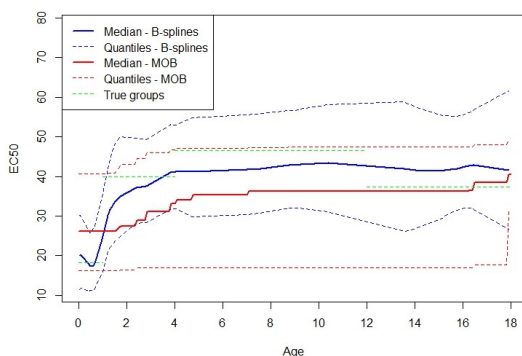
estimates provided by Marshall and Kearns³⁷, and we assume a normally distributed random error with mean zero and variance 15. We restrict attention to a hyperbolic Emax model because estimates of age group specific Hill parameters are not reported by Marshall and Kearns. Using these simulated data, we fitted a two parameter Emax model separately to each age group. Figure 8 shows the four fitted curves.

8.1 Bayesian penalised B-splines

We implement the Bayesian penalised B-splines model by running three Markov chains using a thinning rate of 3 and 9000 iterations, 4500 of which are discarded as burn-in samples. We adopt the first-order random walk prior defined in Section 4.3 for the penalisation. We found a great deal of sensitivity, in terms of convergence, to the choice of prior for the standard deviation parameters of the random walk priors on the B-spline coefficients of the Emax and EC50 parameters. This sensitivity was found when using the Inverse-Gamma priors as used in Section 6. We would advise caution and appropriate checks to ensure posterior results are reliable. One should check a priori the plausible range of values for these standard deviations, which would depend on the magnitude of the Emax and EC50 parameters. Gamma(2, 1/A) priors, with A large (such as $A = 10$) are recommended by Chung et al. (2013)³⁸ and the Stan user guide³⁹ as boundary-avoiding priors in hierarchical models for hierarchical standard deviations. Placing Gamma(2, 0.1) priors on the random walk prior standard deviations allowed the two parameter Emax model to fit well to the simulated



(a) Emax



(b) EC50

Figure 9. Plots of the Bayesian penalised B-spline and bootstrapped MOB fits of (a) the Emax parameter and (b) the EC50 parameter. The median of each parameter, with 2.5th and 97.5th quantiles, over the 1000 simulated bootstrap samples and true parameter values reported by Marshall and Kearns³⁷ given by the green dotted lines are also shown.

data shown in Figure 8, with the chains converging with Gelman-Rubin convergence diagnostic < 1.011 for all parameters.

Figure 9 shows the fitted Bayesian penalised B-spline for the Emax and EC50 parameters over age, showing the median, 2.5th and 97.5th quantiles. The fitted B-splines for both the EC50 and Emax parameters seem to follow closely to the true underlying parameter values and, as can be seen from Figure 10a, the underlying E-R relationships are accurately estimated. Figure 10a plots fitted expected response against concentration in each of the four age groups. The fitted expected response is calculated by setting the Emax and EC50 to values obtained by evaluating the Emax and EC50 fitted B-splines at the mid-points of each age group.

8.2 Bootstrapped MOB

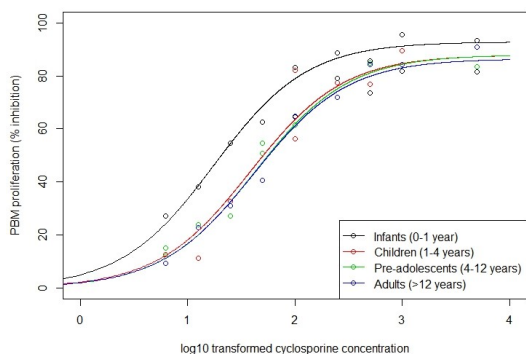
To implement the bootstrapped MOB approach, we used the ‘mob’ function in R with a two parameter Emax model. Otherwise, the approach proceeds exactly as the bootstrapped PALM trees approach described in Section 4.2. To incorporate a two parameter Emax model in the ‘mob’ function, we built on code provided by Thomas and Bornkamp¹⁴, using the ‘nls’ function in R²⁶ to specify the two parameter Emax model.

Figure 9 shows the fitted bootstrapped MOB for the Emax and EC50 parameters over age, showing the median, 2.5th and 97.5th quantiles over the bootstrapped samples. The fitted Emax and EC50 parameters do change with age. However, they are both quite far from the true underlying values. When looking at Figure 10b we see that the model still fits fairly well to the general shape of the data. However, Figure 10b highlights that there is worse separation between the fitted E-R curves for different age groups across the whole concentration range when using the bootstrapped MOB approach as compared with the Bayesian penalised B-splines.

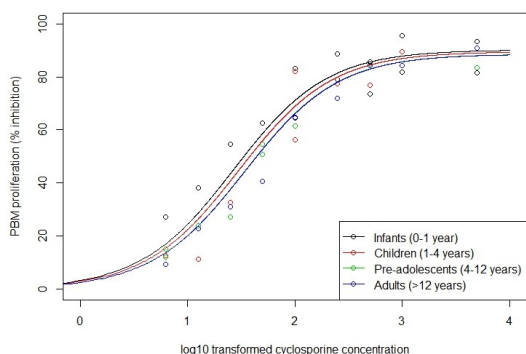
8.3 Deriving dosing rules

Following the procedure to derive optimal dosing rules described in Section 5, Figure 11 provides a plot of the objective function values for rules based on both the Bayesian penalised B-splines and bootstrapped MOB approaches. Overall, the bootstrapped MOB approach achieves lower objective function values. For both approaches, two age groups would almost certainly be recommended by visual inspection.

For two age groups, the optimal age groups defining the bootstrapped MOB dosing rule would be 0 to 3.33 years and 3.33 to 18 years, with target exposures of 191.95 and 294.87, respectively. The optimal age groups defining the Bayesian penalised B-splines dosing rule would be 0 to 0.84 years and 0.84 to 18 years, with target exposures of 110.36 and 446.04, respectively. It is interesting to note how different the dosing rules are for these two methods: the bootstrapped MOB rule stipulates a wider youngest age group, with larger target exposure levels than the Bayesian penalised B-splines rule. However, overall the bootstrapped MOB dosing rule has a lower maximum target exposure than the Bayesian penalised B-splines dosing rule. This seems



(a) B-splines



(b) MOB

Figure 10. Fitted relationships between log base-10 transformed cyclosporine concentrations and PBM proliferation based on parameter estimates for the four age groups obtained with (a) the Bayesian penalised B-spline approach and (b) the bootstrapped MOB approach.

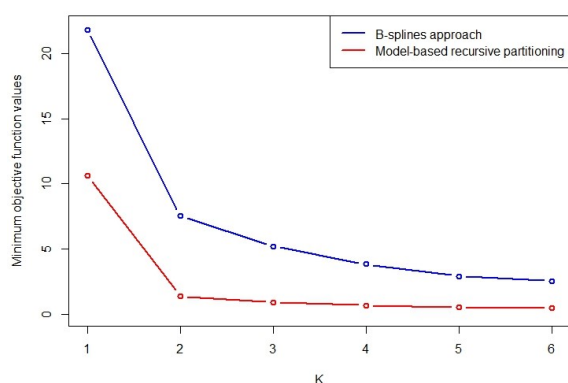


Figure 11. Plot of the objective function values from the optimisation procedure used to identify age groups for the Bayesian penalised B-splines approach (blue line) and bootstrapped MOB (red line).

to be indicative of the larger differences between the age group specific E-R relationships found using the Bayesian penalised B-splines approach.

9 Discussion

In this paper we have considered several approaches to estimating if and how E-R model parameters change with age in order to determine practical dosing rules for distinct paediatric age groups. Our approaches concentrate on the relationship between exposure and response, deriving target exposures for age groups. These target exposures can then be used to identify dosing rules based on a separate relationship between dose and exposure. We do not develop PK models relating dose and exposure in this paper, although many methods exist to do this⁴⁰. In other therapeutic areas, non-monotonic changes in E-R model parameters over some age intervals may be plausible. Evaluating the performance of our methods in these scenarios is outside the scope of this paper, but is something that could be investigated in future research.

We derive the target exposures for each age group by taking the age group mid-point and finding the exposure level at which the expected response would be equal to the target response. In reality, this may not actually be the optimal exposure level over the whole age group. Rather than using the exposure appropriate for the age group midpoint, an alternative approach for deriving a more accurate target exposure would be the following: within an age group, target the single exposure level which minimises the total absolute difference between the expected response associated with the target exposure and the target response, integrating over the age group. This approach is computationally more demanding making it unsuitable for use in our simulation study, but can be quickly implemented for one dataset in practice.

Results of our simulations with linear E-R models suggest that the Bayesian penalised B-splines and bootstrapped PALM tree approaches perform similarly in terms of estimating changes in E-R model parameters over age, though the integrated absolute bias and empirical SD is consistently lower in the Bayesian penalised B-splines approach. Plots of the absolute difference between the true expected response implied by proposed target exposures and the target response also suggest that for most scenarios both approaches perform similarly, though in some scenarios Bayesian penalised B-splines perform better than bootstrapped PALM trees, and vice versa. In fact, the Bayesian

penalised B-splines approach appears to outperform all other approaches in most scenarios; only the approach using categorical covariates sometimes has lower integrated absolute bias, and even then, only in scenarios where the true underlying E-R models contain four age groups matching ICH E11 guidance (as is assumed in the categorical covariates approach).

10 Acknowledgments

IW and LVH are grateful to acknowledge funding from the UK Medical Research Council (MR/M013510/1). IW's contribution to the manuscript was made while he was an employee of Lancaster University. This work is independent research and Professor Jaki's contribution to it was funded by his Senior Research Fellowship (NIHR-SRF-2015-08-001) supported by the National Institute for Health Research. The views expressed in this publication are those of the authors and not necessarily those of the UK Medical Research Council or the National Institute for Health Research. We are grateful to Dr. Graeme Sills of the University of Liverpool for his advice on realistic simulation scenarios and target responses, and to Marius Thomas for sharing R code to fit MOB with more complex models.

References

1. ICH. E11: Note for guidance on clinical investigation of medicinal products in the paediatric population, 2001. Accessed: 29/12/2017. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002926.pdf.
2. Hampson LV, Herold R, Posch M et al. Bridging the gap: A review of dose investigations in paediatric investigation plans. *British Journal of Clinical Pharmacology* 2014; 78(4): 898–907.
3. Takahashi H, Ishikawa S, Nomoto S et al. Developmental changes in pharmacokinetics and pharmacodynamics of warfarin enantiomers in Japanese children. *Clinical Pharmacology & Therapeutics* 2000; 68(5): 541–555.
4. Dunne J, Rodriguez WJ, Murphy MD et al. Extrapolation of adult data and other data in pediatric drug-development programs. *Pediatrics* 2011; 128(5): 1242–1249.
5. Macdougall J. *Analysis of Dose-Response Studies—Emax Model*, chapter 9. Springer New York, 2006. pp. 127–145. DOI:10.1007/0-387-33706-7_9.
6. Thomas N, Sweeney K and Somayaji V. Meta-analysis of clinical dose-response in a large drug development portfolio. *Statistics in Biopharmaceutical Research* 2014; 6(4): 302–317.
7. Parkinson J, Tang W, Johansson CC et al. Comparison of the exposure-response relationship of dapagliflozin in adult and paediatric patients with type 2 diabetes mellitus. *Diabetes, Obesity and Metabolism* 2016; 18(7): 685–692.
8. Earp JC, Mehrotra N, Peters KE et al. Esomeprazole FDA approval in children with GERD: Exposure-matching and exposure-response. *Journal of Pediatric Gastroenterology and Nutrition* 2017; 65(3): 272–277.
9. Lipkovich I, Dmitrienko A et al. Tutorial in biostatistics: Data-driven subgroup identification and analysis in clinical trials. *Statistics in Medicine* 2017; 36(1): 136–196.
10. Ondra T, Dmitrienko A, Friede T et al. Methods for identification and confirmation of targeted subgroups in clinical trials: a systematic review. *Journal of Biopharmaceutical Statistics* 2016; 26(1): 99–119.
11. Eilers PH and Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science* 1996; 11(2): 89–102.
12. Seibold H, Hothorn T and Zeileis A. Generalised linear model trees with global additive effects. *arXiv preprint arXiv:161207498* 2016; .
13. Zeileis A, Hothorn T and Hornik K. Model-based recursive partitioning. *Journal of Computational and Graphical Statistics* 2008; 17(2): 492–514.
14. Thomas M, Bornkamp B and Seibold H. Subgroup identification in dose-finding trials via model-based recursive partitioning. *Statistics in Medicine* 2017; : 1–17DOI:10.1002/sim.7594. Available online as Early Access.
15. ICH. E4: Dose-response information to support drug registration, 1994. Accessed: 26/01/2018. Available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E4/Step4/E4_Guideline.pdf.
16. Berges A and Chen C. Dose finding by concentration-response versus dose-response: A simulation-based comparison. *European Journal of Clinical Pharmacology* 2013; 69(7): 1391–1399.
17. Gibaldi M and Perrier D. *Pharmacokinetics*. Drugs and the Pharmaceutical Sciences Series, Marcel Dekker, 1975.
18. Mould D and Upton RN. Basic concepts in population modeling, simulation, and model-based drug development—part 2: introduction to pharmacokinetic modeling methods. *CPT: pharmacometrics & systems*

- pharmacology* 2013; 2(4): 1–14.
19. Hsu CH. Evaluating potential benefits of dose–exposure–response modeling for dose finding. *Pharmaceutical Statistics* 2009; 8(3): 203–215.
 20. Schoemaker R, Wade JR and Stockis A. Brivaracetam population pharmacokinetics and exposure–response modeling in adult subjects with partial-onset seizures. *The Journal of Clinical Pharmacology* 2016; 56(12): 1591–1602.
 21. Girgis IG, Nandy P, Nye JS et al. Pharmacokinetic–pharmacodynamic assessment of topiramate dosing regimens for children with epilepsy 2 to < 10 years of age. *Epilepsia* 2010; 51(10): 1954–1962.
 22. Nedelman JR, Rubin DB and Sheiner LB. Diagnostics for confounding in PK/PD models for oxcarbazepine. *Statistics in Medicine* 2007; 26(2): 290–308.
 23. Kallen A. *Computational pharmacokinetics*. Chapman and Hall/CRC, 2007.
 24. Zeileis A and Hornik K. Generalized m-fluctuation tests for parameter instability. *Statistica Neerlandica* 2007; 61(4): 488–508.
 25. Hothorn T and Zeileis A. Partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research* 2015; 16: 3905–3909.
 26. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL <https://www.R-project.org/>.
 27. Breiman L. Bagging predictors. *Machine Learning* 1996; 24(2): 123–140.
 28. Bowman A and Evers L. Nonparametric smoothing, 2016. APTS Lecture notes. Accessed: 07/02/2018. Available at: <http://www2.warwick.ac.uk/fac/sci/statistics/apts/students/resources-1617/smoothing-notes.pdf>.
 29. Lang S and Brezger A. Bayesian P-splines. *Journal of Computational and Graphical Statistics* 2004; 13(1): 183–212.
 30. Kneib T, Brezger A and Crainiceanu CM. *Generalized Semiparametric Regression with Covariates Measured with Error*. In: Kneib T, Tutz G (eds) *Statistical Modelling and Regression Structures - Festschrift in Honour of Ludwig Fahrmeir*. Springer, 2010.
 31. Carpenter B, Gelman A, Hoffman M et al. Stan: A probabilistic programming language. *Journal of Statistical Software* 2016; 20(2): 1–37.
 32. Stan Development Team. RStan: the R interface to Stan, 2016. URL <http://mc-stan.org/>. R package version 2.14.1.
 33. Williams K, Thomson D, Seto I et al. Standard 6: Age groups for pediatric trials. *Pediatrics* 2012; 129(Supplement 3): S153–S160.
 34. Wadsworth I, Hampson LV, Jaki T et al. A quantitative framework to inform extrapolation decisions in children. *Under review* 2018; .
 35. Jennison C and Turnbull BW. *Group Sequential Methods with Applications to Clinical Trials*. CRC Press, 1999.
 36. Burden RL and Faires JD. *Numerical Analysis*. Cengage Learning, 2011.
 37. Marshall JD and Kearns GL. Developmental pharmacodynamics of cyclosporine. *Clinical Pharmacology & Therapeutics* 1999; 66(1): 66–75.
 38. Chung Y, Rabe-Hesketh S, Dorie V et al. A non-degenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika* 2013; 78(4): 685–709.
 39. Stan Development Team. Stan modeling language users guide and reference manual, version 2.17.0, 2017. URL <http://mc-stan.org/>.
 40. Rowland M and Tozer TN. *Clinical Pharmacokinetics and Pharmacodynamics: Concepts and Applications*. Lippincott Williams and Wilkins, 2010.

11 Appendix A: Simulation scenarios in detail

In terms of how the E-R model parameters change over age, we consider 11 scenarios for data generation:

1. Step function relates how E-R model parameters change over age, following ICH guidance document age groupings with substantial differences between E-R model parameter values in each age group;
2. E-R model parameters have a less steep linear transition between age groups, with parameter values and age groups the same as scenario 1;
3. E-R model parameter values have same change over age as in scenario 2, though now age groups parameter values are more similar between age groups to more closely resemble what might be observed in reality;
4. Step function relates how E-R model parameters change over age, as in scenario 1, but now there is a deviation from the ICH age groups so that there

are more distinct age groups amongst younger children, following NICHD groupings;

5. E-R model parameters constant over all age groups;
6. E-R model parameters have a constant linear decrease over the whole age range;
7. Intercept term as in scenario 3, slope term constant over age as in scenario 6;
8. Intercept term constant over age as in scenario 6, slope term as in scenario 3;
9. Same as scenario 3, though now the true age groups will be 0 to 6 months, 6 months to 3 years, 3 years to 11 years, 11 years to 18 years;
10. Same as scenario 3, though now the true age groups will be 0 to 2 year, 2 year to 6 years, 6 years to 14 years, 14 years to 18 years;
11. Same as scenario 3, though now the true age groups will be 0 to 8 years, 8 years to 12 years, 12 years to 16 years, 16 years to 18 years.