

J. R. Statist. Soc. A (2020)
183, Part 2, pp. 515–534

A quantitative framework to inform extrapolation decisions in children

Ian Wadsworth,

Lancaster University and Phastar, Macclesfield, UK

Lisa V. Hampson,

Novartis Pharma, Basel, Switzerland

Thomas Jaki,

Lancaster University, UK

Graeme J. Sills,

University of Liverpool, UK

Anthony G. Marson

University of Liverpool and Walton Centre National Health Service Foundation Trust, Liverpool, UK

and Richard Appleton

Alder Hey Children's Hospital, Liverpool, UK

[Received December 2017. Final revision September 2019]

Summary. When developing a new medicine for children, the potential to extrapolate from adult efficacy data is well recognized. However, significant assumptions about the similarity of adults and children are needed for extrapolations to be biologically plausible. One such assumption is that of similar exposure–response (E – R -) relationships. Motivated by applications to antiepileptic drug development, we consider how data that are available from existing trials of adults and adolescents can be used to quantify prior uncertainty about whether E – R -relationships are similar in adults and younger children. A Bayesian multivariate meta-analytic model is fitted to existing E – R -data and adjusted for external biases that arise because these data are not perfectly relevant to the comparison of interest. We propose a strategy for eliciting expert prior opinion on external biases. From the bias-adjusted meta-analysis, we derive prior distributions quantifying our uncertainty about the degree of similarity between E – R -relationships for adults and younger children. Using these we calculate the prior probability that average pharmacodynamic responses in adults and younger children, both on placebo and at an effective concentration, are sufficiently similar to justify a complete extrapolation of efficacy data. A simulation study is performed to evaluate the operating characteristics of the approach proposed.

Keywords: Bias; Clinical trials; Elicitation; Extrapolation; Meta-analysis; Paediatrics

Address for correspondence: Lisa V. Hampson, Advanced Methodology and Data Science, Novartis Pharma AG, Postfach 4002, Basel, Switzerland.
E-mail: lisa.hampson@novartis.com

© 2019 The Authors. Journal of the Royal Statistical Society: Series A (Statistics in Society) 0964–1998/20/183515 published by John Wiley & Sons Ltd on behalf of the Royal Statistical Society.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Leveraging existing data to optimize the design of a drug development programme is particularly appropriate when we develop medicines for small or vulnerable populations, such as children. The European Medicines Agency (EMA) defines extrapolation as

‘... extending information and conclusions available from studies in one or more subgroups of the patient population (source population)... to make inferences for another subgroup of the population (target population)...’

(European Medicines Agency, 2013, 2017). We focus on the extrapolation of adult efficacy data to children. Wadsworth *et al.* (2016a) reported the findings of a systematic review of statistical methods that are relevant for extrapolating efficacy and other data from adults to children. Gamalo-Siebers *et al.* (2017) illustrated methods that use adult data to design improved paediatric programmes and Petit *et al.* (2018) proposed a method for the design and analysis of paediatric dose-finding trials, with the dose range calculated by extrapolating from adult pharmacokinetic data. Weber *et al.* (2018) compared the use of Bayesian and frequentist methods for combining existing adult and paediatric data to inform decision making. Crippa *et al.* (2018) developed a one-stage approach to the meta-analysis of aggregated data which can be used for estimating non-linear dose–response models. Beyond extrapolation from adult data, Zheng *et al.* (2019) proposed a Bayesian hierarchical model synthesizing animal and human toxicity data to learn about the relationship between dose and toxicity risk in a phase I human oncology trial.

To justify the extrapolation of adult efficacy data to children, we must often make strong assumptions about the similarity of age groups in terms of disease progression, response to intervention and exposure–response (*E–R*-) relationships which we take to link a single outcome measure to a single summary of exposure. These assumptions are made explicit in the paediatric decision tree (see Food and Drug Administration (2003)) where judgements about the plausibility of each aspect of similarity determine whether, in the terminology of Dunne *et al.* (2011), a ‘complete’, ‘partial’ or ‘no’ extrapolation strategy is adopted. Fig. 1 explains the implications of the extrapolation strategy for the data that are generated in paediatrics. Safety data are required regardless of the extrapolation strategy that is adopted. Dunne *et al.* (2011) reviewed 370 paediatric studies submitted to the US Food and Drug Administration between 1998 and 2008 to identify cases in which efficacy data were extrapolated: of the 166 drug products that were considered, 14.5% followed a complete extrapolation strategy, 68% a partial extrapolation strategy and 17.5% did not extrapolate. Sun *et al.* (2017), in an update to this review, considered 388 paediatric studies that were submitted between 2009 and 2014. The proportion of products using partial extrapolation fell to 29%, whereas the proportions using no and complete extrapolation rose to 37% and 34% respectively.

Since 2006, the European Union paediatric regulation (European Union, 2006) has mandated that studies that are intended to support licensing of a medicine for children in the European Union must follow a paediatric investigation plan, which must be agreed ahead of time with the European Medicines Agency’s Paediatric Committee. When selecting an extrapolation strategy, sponsors must ask themselves how plausible assumptions required for extrapolation are, given the data to hand. Hlavín *et al.* (2016) used a scepticism factor to represent uncertainty about the plausibility of complete extrapolation, where this factor could be established from existing data or expert opinion. This paper presents a framework using existing data to inform a decision on whether to perform a complete or partial extrapolation of efficacy data from adults to children. This decision will determine whether the sponsor will collect only pharmacokinetic data in children to support dose finding, or both pharmacokinetic and pharmacodynamic data.

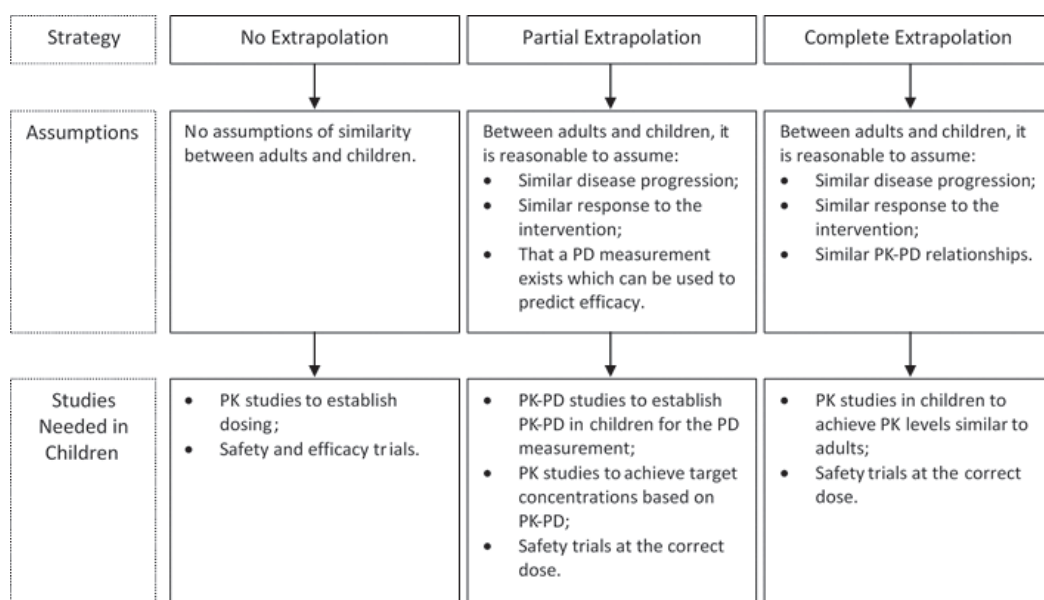


Fig. 1. Extrapolation strategies, assumptions made and required studies, based on the Food and Drug Administration paediatric decision tree (Food and Drug Administration, 2003) (PK, pharmacokinetic; PD, pharmacodynamic)

The framework proposed begins with sponsors prespecifying numerical criteria which E – R curves in adults and children must satisfy to be deemed ‘similar’. The sponsor can then use existing information to quantify prior confidence in this degree of similarity.

We propose performing a Bayesian random-effects meta-analysis of existing E – R data to derive priors for differences between E – R relationships in adults and children. When studying small populations it is likely that few existing studies will be available for synthesis. Furthermore, ‘external biases’ (Turner *et al.*, 2009) may be inherent in the existing data if there are differences between the source and target populations, e.g. if existing data are measurements on adults and adolescents but our question is whether E – R relationships in adults and children aged 2–11 years are similar. This scenario may often arise in practice because drug development in adults and children is typically staggered. Furthermore, adolescents are also often recruited into adult trials in therapeutic areas such as epilepsy (Girgis *et al.*, 2010; French *et al.*, 2012; Marson *et al.*, 2007a,b) and asthma (FitzGerald *et al.*, 2016; O’Byrne *et al.*, 2017). A draft guidance document for the inclusion of adolescents in adult oncology trials has recently been released by the Food and Drug Administration (Food and Drug Administration, 2018). To derive prior distributions for key parameters accounting for external biases, existing data could be downweighted according to a fixed weight (e.g. Ibrahim and Chen (2000), Tan *et al.* (2003) and Rietbergen *et al.* (2011)) or dynamically downweighted to a degree reflecting the commensurability of the new and existing data (Ibrahim and Chen, 2000; Hobbs *et al.*, 2011; Neuenschwander *et al.*, 2010). The challenges of dynamic downweighting were noted in Galwey (2017). Alternatively, one could model the external biases and define either empirical priors for the bias parameters (Welton *et al.*, 2009) or elicit expert opinion on them (Turner *et al.*, 2009). We adopt the latter approach here.

Throughout we illustrate the proposed extrapolation framework with applications to anti-epileptic drug development. In this setting, there is broad agreement about the acceptability of

extrapolating efficacy data in adults with partial onset seizures to older children with partial onset seizures, although there is some uncertainty about what age one can extrapolate down to (European Medicines Agency, 2010; Pediatric News, 2016; French *et al.*, 2004; Wadsworth *et al.*, 2016b). This paper proceeds as follows. In Section 2 we define a Bayesian bias-adjusted multivariate meta-analytic model to synthesize existing E – R -data and propose a quantitative criterion for defining similar E – R -relationships in adults and younger children. Section 3 describes a scheme for eliciting expert opinion on the external biases that may be inherent in the existing data. In Sections 4 and 5, we report a simulation study that was used to evaluate properties of our framework before concluding in Section 6 with a discussion.

2. Using existing data to inform an extrapolation decision

2.1. Motivation

Suppose that E – R -data are available from H existing trials which recruited adults and adolescents. Let Y_{ij} represent the response of the i th subject in study j , for $i = 1, \dots, N_j$, and $j = 1, \dots, H$. Dropping the i - and j -subscripts for clarity, let A be a binary indicator of age which takes the value 1 for adolescents and 0 otherwise. Let

$$\mathbb{E}(Y) = g^{-1} \left(\gamma_0 + \sum_{k=1}^K \gamma_k x_k + \gamma_C C + \gamma_A A + \gamma_I CA \right), \quad (1)$$

where C is a measure of drug exposure, x_1, \dots, x_K are baseline covariates (such as weight) influencing response and g is the link function of the generalized linear model.

If we can assume that regression parameters remain constant across studies, the relationship between exposure and the expected pharmacodynamic response, hereafter referred to as the E – R -‘relationship’ or ‘curve’, is identical in adults and adolescents if $\gamma_A = \gamma_I = 0$. The assumption of between-trial homogeneity is relaxed in Section 2.3, in which case γ_A and γ_I are interpreted as mean parameters.

To simplify the presentation of our methods, we shall assume throughout that the pharmacodynamic response of interest is normally distributed and that a generalized linear model is an adequate description of the underlying E – R -relationship:

$$Y = \gamma_0 + \sum_{k=1}^K \gamma_k x_k + \gamma_C C + \gamma_A A + \gamma_I CA + \epsilon \quad (2)$$

where $\epsilon \sim N(0, \sigma^2)$ is a random error term. Linear models have been used to analyse E – R -data for the antiepileptic drugs oxcarbazepine (Nedelman *et al.*, 2007) and topiramate (Girgis *et al.*, 2010) setting $Y = \log(Z + 110)$, where Z is the percentage change from baseline in seizure frequency and C represents the steady state trough concentration under repeated dosing (C_{\min}).

Consider now the data that we would accumulate if we performed an E – R -study, indexed by T , in adults and younger children. The International Conference on Harmonisation E11 guidance (International Conference on Harmonisation, 2001) defines children as aged 2–11 years, and adolescents as 12–16 or 18 years. If we made a complete extrapolation of efficacy data from adults to younger children, we would not need to perform study T but it is useful to consider the data that it would generate. Suppose that we measure pharmacodynamic responses Y_{iT} , for $i = 1, \dots, N_T$. Again, dropping the subscript i for clarity, let

$$Y_T = \beta_0 + \sum_{k=1}^K \beta_k x_{kT} + \beta_C C_T + \beta_A A_T + \beta_I C_T A_T + \epsilon_T, \quad (3)$$

where $\epsilon_T \sim N(0, \sigma^2)$, x_{1T}, \dots, x_{KT} are baseline prognostic covariates defined analogously to x_1, \dots, x_K , C_T is a measure of exposure defined similarly to C and A_T is a binary age covariate taking the value 1 for younger children and 0 otherwise. E – R –relationships in adults and younger children are identical if $\beta_A = \beta_I = 0$.

We relate parameters in the source and target populations described by models (2) and (3) as

$$\begin{aligned}\beta_A &= \gamma_A + \delta_A, \\ \beta_I &= \gamma_I + \delta_I.\end{aligned}\tag{4}$$

Here δ_A and δ_I represent external biases arising because E – R –curves in adolescents and younger children may differ because of the effects of maturation and physical development on drug absorption, distribution, metabolism and elimination, and on the action of and response to a drug (Kearns *et al.*, 2003). Stephenson (2005) noted that the responses of adults and children to many drugs have much in common, although there are exceptions, such as warfarin (Takahashi *et al.*, 2000) and cyclosporine (Marshall and Kearns, 1999). An alternative to the additive bias model (4) is a proportional model stipulating $\beta_A = \delta_A \gamma_A$ and $\beta_I = \delta_I \gamma_I$ (Turner *et al.*, 2009). We prefer an additive model since there may be differences between adults and younger children even if no differences between adults and adolescents exist. The existing data \mathcal{D}_E are said to be relevant for learning about likely differences between E – R –relationships in adults and younger children if δ_A and δ_I are both close to 0.

2.2. Extrapolation criterion

We propose criteria evaluating whether a summary measure of the distribution of pharmacodynamic responses in adults and younger children on placebo and at an effective exposure are sufficiently similar. Let C^* denote a level of exposure that is known to be effective in adults, e.g. the adult EC_{90} , the exposure at which the expected adult response is 90% of the maximum. It may be more straightforward to specify equivalence margins with differences between a transformed outcome in mind. Thus, E – R –curves are said to be similar if

$$\mathcal{M}\{h(Y_T|C_T=0; A_T=1)\} - \mathcal{M}\{h(Y_T|C_T=0; A_T=0)\} \in (-\eta_1, \eta_1)\tag{5}$$

and

$$\mathcal{M}\{h(Y_T|C_T=C^*; A_T=1)\} - \mathcal{M}\{h(Y_T|C_T=C^*; A_T=0)\} \in (-\eta_2, \eta_2),\tag{6}$$

where $h(Y_T|C_T, A_T)$ is a function of the pharmacodynamic response of a subject with observed exposure C_T in age group A_T , and \mathcal{M} is a measure of location such as the mean or median. We require the distribution of pharmacodynamic outcomes in adults and younger children to be similar at the adult effective concentration. This is because, if E – R –relationships are judged to be similar between these two age groups, a suitable dose for children would be found by matching exposures. In practice, bounds η_1 and η_2 would be set on the basis of clinical judgement. Larger bounds imply that larger differences between the average pharmacodynamic responses of adults and younger children will be tolerated if we incorrectly perform a complete extrapolation and dose younger children targeting the adult effective concentration. Although different equivalence bounds can be applied at $C_T=0$ and $C_T=C^*$, to simplify we set $\eta_1 = \eta_2$.

The joint prior probability of extrapolation criteria (5)–(6), denoted by p_E , can be used to measure the prior plausibility of an assumption that E – R –curves are sufficiently similar in adults and children to justify a complete extrapolation of efficacy data across these age groups. We speculate that p_E in excess of 0.8 or 0.9 would be sufficient to support the immediate adoption of a

complete extrapolation strategy. Lower probabilities would prompt a sponsor to collect additional E – R -data in younger children to verify similarities and to facilitate dose finding, where the exact sample size could be determined according to an expected value of information calculation (Willan and Pinto, 2005; Wilson, 2015). A very low value of p_E could be consistent either with extreme uncertainty about the relevance of the existing data or a strong degree of scepticism about the similarity of E – R -curves. In both cases, the most appropriate strategy would be to plan an E – R -study in younger children sized to support independent dose finding in this age group.

2.3. Bayesian bias-adjusted meta-analytic model for existing data

We begin the process of quantifying what is known about differences between adults and younger children by performing a Bayesian meta-analysis of adult and adolescent E – R -data to learn about γ_A and γ_I . We assume that individual patient data are available but aggregate data could be used if maximum likelihood estimates and their standard errors are available for all parameters in the linear predictor of model (2). At the first level of the model, data from study j , $j = 1, \dots, H$, enrolling adults and adolescents are modelled as

$$Y_{ij} = \gamma_{0j} + \gamma_{Cj}C_{ij} + \gamma_{Aj}A_{ij} + \gamma_{Ij}C_{ij}A_{ij} + \epsilon_{ij}, \quad (7)$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$ and for ease of presentation we assume that the only baseline covariate prognostic for outcome is age. To limit model complexity, we regard $\gamma_{01}, \dots, \gamma_{0H}$ and $\gamma_{C1}, \dots, \gamma_{CH}$ as study-specific intercepts and effects of exposure respectively and make no assumption about exchangeability.

For the remaining parameters in model (7), we assume that pairs of study-specific parameters $(\gamma_{A1}, \gamma_{I1}), \dots, (\gamma_{AH}, \gamma_{IH})$ are exchangeable and are samples from a bivariate normal random-effects distribution with mean $\mu = (\gamma_A, \gamma_I)$ and covariance matrix Σ . One approach would be to place an inverse Wishart prior on Σ . However, our investigations found that the results of meta-analyses are very sensitive to the choice of the inverse Wishart scale matrix; decreasing the diagonal elements of this matrix reduces the variances of the marginal posterior distributions of γ_A and γ_I . Gelman (2006) showed that inverse-gamma(ε, ε) priors with $\varepsilon \approx 0$ are informative for variance parameters in hierarchical models and suggested that inverse Wishart prior distributions for covariance matrices incur similar issues. To avoid this sensitivity, we adopt an alternative parameterization (Medical Research Council Biostatistics Unit, 2017) for the bivariate normal random-effects distribution which gives the analyst more flexibility in how they specify priors for the variance parameters. For $j = 1, \dots, H$, we define

$$\begin{aligned} \gamma_{Aj} &\sim N(\gamma_A, \xi_1^2), \\ \gamma_{Ij} | \gamma_{Aj} &\sim N\{\lambda_0 + \lambda_1(\gamma_{Aj} - \bar{\gamma}_A), \xi_2^2\}, \end{aligned}$$

where $\bar{\gamma}_A = (1/H)\Sigma\gamma_{Aj}$, which implies that

$$\begin{pmatrix} \gamma_{Aj} \\ \gamma_{Ij} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \gamma_A \\ \gamma_I \end{pmatrix}, \begin{pmatrix} \xi_1^2 & \lambda_1 \xi_1^2 \\ \lambda_1 \xi_1^2 & \xi_2^2 + \xi_1^2 \lambda_1^2 \end{pmatrix} \right\}. \quad (8)$$

Under this representation, $\gamma_{Ij} = \lambda_0 + \lambda_1(\gamma_{Aj} - \bar{\gamma}_A)$ and we allow for a correlation between γ_{Aj} and γ_{Ij} , for each $j = 1, \dots, H$.

The meta-analytic model is completed by defining priors for all unknown parameters. For each j , $j = 1, \dots, H$, the study-specific intercept and effect of exposure, γ_{0j} and γ_{Cj} , are assigned independent $N(0, \zeta^2)$ priors. We define average parameters γ_0 and γ_C as $\Sigma_{j=1}^H \gamma_{0j}/H$ and

$\Sigma_{j=1}^H \gamma_{Cj}/H$ so that they represent means for the intercept and effect of exposure across the H existing studies. For the residual precision we stipulate $\sigma^{-2} \sim \text{gamma}(a, b)$, with a and b chosen to define a weakly informative prior. For the parameters of random-effects distribution (8), we place an $N(0, 100)$ prior on γ_A and specify priors $\xi_1 \sim \text{gamma}(a_1, b_1)$, $\xi_2 \sim \text{gamma}(a_2, b_2)$, $\lambda_0 \sim \text{Student-}t(\mu_{\lambda_0}, \sigma_{\lambda_0}, \nu_{\lambda_0})$ and $\lambda_1 \sim N(\mu_{\lambda_1}, \sigma_{\lambda_1}^2)$. In the examples that we have considered, we have chosen hyperparameters to ensure that the prior for the correlation between each pair $(\gamma_{Aj}, \gamma_{Ij})$ has a bucket shape, placing probability mass at -1 and 1 , and furthermore that prior probability mass is placed on a range of plausible values for the between-trial standard deviations of the γ_{Aj} - and γ_{Ij} -parameters (Neuenschwander *et al.*, 2010).

The Bayesian meta-analytic model can be fitted by using Markov chain Monte Carlo (MCMC) sampling. The joint posterior distribution of $(\gamma_0, \gamma_C, \gamma_A, \gamma_I)$ will not be of a standard form. To justify a complete extrapolation decision, the prior probability of criteria (5) and (6) would probably need to be reported in the paediatric investigation plan, trial protocols and journal publications. Using an approximation to the joint posterior distribution which has a closed form would allow reviewers of these documents to reproduce p_E more easily. Otherwise, one would need to rerun the original meta-analysis to generate p_E , which would require access to subject level data which may not be publicly available. Therefore, to facilitate communication and reproducibility of the joint posterior, we approximate it as a finite mixture of K four-dimensional multivariate normal distributions (Schmidli *et al.*, 2014) using the `flexmix` package (Leisch, 2004; Grün and Leisch, 2007, 2008) in R (R Development Core Team, 2015):

$$f(\gamma_0, \gamma_C, \gamma_A, \gamma_I | \mathbf{Y}_1, \dots, \mathbf{Y}_H) \approx \sum_{i=1}^K \omega_i \phi_4(\boldsymbol{\mu}_i, \Sigma_i), \quad (9)$$

where $\phi_4(\boldsymbol{\mu}, \Sigma)$ is the four-dimensional multivariate normal probability density function with mean $\boldsymbol{\mu}$ and variance Σ , and $\mathbf{Y}_1, \dots, \mathbf{Y}_H$ are vectors representing the adult and adolescent data from existing studies $1, \dots, H$. Increasing K in approximation (9) increases the accuracy of the finite mixture approximation as measured by the Kullback–Leibler divergence (Kullback and Leibler, 1951; Schmidli *et al.*, 2014). However, these increases diminish with K and must be balanced against increases in model complexity. In our investigations, we have found that setting $K = 2$ in approximation (9) is adequate.

If we consider (γ_A, γ_I) in model (2) to be systematically biased for the parameters (β_A, β_I) in model (3), then we can elicit expert opinion on the size of these external biases. We assume that prior opinion on the bias parameters can be modelled as a bivariate normal distribution, written as $\boldsymbol{\delta} \sim N_2(\boldsymbol{\nu}, \Pi)$, where $\boldsymbol{\delta} = (\delta_A, \delta_I)$. By sampling pairs (γ_A, γ_I) and (δ_A, δ_I) from $f(\gamma_0, \gamma_C, \gamma_A, \gamma_I | \mathbf{Y}_1, \dots, \mathbf{Y}_H)$ and $\phi_2(\boldsymbol{\nu}, \Pi)$ respectively, we can generate samples from the prior distribution of (β_A, β_I) given the existing data. Fitting these Monte Carlo samples by using maximum likelihood estimation, we obtain the approximate prior

$$g(\beta_A, \beta_I | \mathbf{Y}_1, \dots, \mathbf{Y}_H) = \sum_{i=1}^2 \omega_i \phi_2(\boldsymbol{\mu}_i, \Sigma_i). \quad (10)$$

3. Eliciting prior opinion on external biases

3.1. Overview

In this section we describe our proposal for eliciting an individual expert's opinion on δ_A and δ_I . We envisage eliciting opinion at a face-to-face meeting of experts by using a two-step process: first elicit the beliefs of individuals; then in a group discussion use behavioural aggregation to obtain a consensus opinion (Hampson *et al.*, 2014). If experts cannot agree on a consensus prior,

mathematical aggregation of individual expert opinion could be considered. Experts should be subject matter specialists, such as consultant level clinicians with a relevant specialism. We recommend identifying such individuals through research groups and networks, trying to achieve a good coverage of potential viewpoints. In this context, we suggest involving clinicians who specialize in treating adults and those who treat paediatrics so that we can draw on the experience of both groups to interpret the adult and adolescent data appropriately. Our elicitation scheme has four components.

- (a) Part 1: present to each expert the fitted dose–response curves for adults and adolescents derived from a meta-analysis of completed trials.
- (b) Part 2: elicit each expert’s prior modal guess at the dose–response curve in younger children, in light of the data that are presented in part 1.
- (c) Part 3: elicit from the expert their uncertainty about their answer to part 2 as a 90% credible interval.
- (d) Part 4: use the expert’s answers to derive a fitted prior for δ_A and δ_I . Feed back summaries of fitted priors for the dose–response relationship in younger children. Allow the expert to revise their answers until they are happy that their fitted prior captures their beliefs.

Note that we frame elicitation questions in terms of the dose–response, rather than E – R –relationship, since clinicians are likely to be more familiar expressing beliefs about the former. In our experience, serum concentrations of antiepileptic (and other drugs) are not typically measured in routine clinical practice, so clinicians tend to be more familiar with dose than with concentration. Answers to elicitation questions can be translated to opinions on E – R –parameters assuming that the relationship between dose and exposure is known. This might be derived by using existing E – R –data or through a further elicitation exercise with pharmacometricians. In our examples, we assume that dose proportionality holds over the dose range of interest. Letting d denote dose, we can write exposure as $C = \kappa d$, with known κ .

3.2. Rationale for the elicitation scheme

We now show that, under certain assumptions, we can deduce, from an expert’s conditional opinions on the dose–response curve in younger children given in parts 2 and 3 of the elicitation scheme, the joint prior distribution for (δ_A, δ_I) .

It is reasonable to suppose that, when presented with the fitted adult and adolescent dose–response curves, the expert will take these to be the true response curves for these age groups, disregarding any estimation error. Ignoring for the moment between-study heterogeneity in E – R –parameters, model (2) stipulates that, at dose d^* , the fitted average pharmacodynamic response in adults is $\mathcal{F}_{d^*}^1 = \gamma_0 + \gamma_C \kappa d^*$ and the fitted average response in adolescents is $\mathcal{F}_{d^*}^1 + \mathcal{F}_{d^*}^2 = \gamma_0 + \gamma_C \kappa d^* + \gamma_A + \gamma_I \kappa d^*$. Therefore, from the presentation of the existing data, at dose d^* an expert can deduce

- (a) the average response in adults, $\mathcal{F}_{d^*}^1$, and
- (b) the difference between the adult and adolescent expected responses $\mathcal{F}_{d^*}^2 = \gamma_A + \gamma_I \kappa d^*$.

Assuming bias model (4) and assuming no drift in the parameters of the adult E – R –relationship so that, in a future E – R –study enrolling adults and younger children $\beta_0 = \gamma_0$ and $\beta_C = \gamma_C$, model (3) stipulates that the expected response of a younger child given dose d^* would be

$$\mathbb{E}(Y_T | A_T = 1, C_T = \kappa d^*) = \gamma_0 + \gamma_C \kappa d^* + \gamma_A + \delta_A + (\gamma_I + \delta_I) \kappa d^*.$$

Conditioning on what has been learnt from the existing data, we have

$$\mathbb{E}(Y_T | A_T = 1, C_T = \kappa d^*) | \mathcal{F}_{d^*}^1, \mathcal{F}_{d^*}^2 = \mathcal{F}_{d^*}^1 + \mathcal{F}_{d^*}^2 + \delta_A + \delta_I \kappa d^*.$$

Assuming that prior opinion on (δ_A, δ_I) is independent of opinion on other $E-R$ model parameters and can be modelled as $N_2(\nu, \Pi)$, with $\nu = (\nu_A, \nu_I)$ and

$$\Pi = \begin{pmatrix} \pi_A^2 & \pi_{AI} \\ \pi_{AI} & \pi_I^2 \end{pmatrix},$$

then at dose d^* we obtain

$$\mathbb{E}(Y_T | A_T = 1, C = \kappa d^*) | \mathcal{F}_{d^*}^1, \mathcal{F}_{d^*}^2 \sim N\{\mathcal{F}_{d^*}^1 + \mathcal{F}_{d^*}^2 + \nu_A + \nu_I \kappa d^*, \pi_A^2 + 2\pi_{AI} \kappa d^* + \pi_I^2 (\kappa d^*)^2\}.$$

In parts 2 and 3 of the elicitation scheme, we ask the expert for their conditional opinion on the average response of younger children on placebo, a ‘medium’ or a ‘high’ dose, denoted by d_0 , d_M and d_H respectively. The proposed wording of the elicitation questions is given in the on-line supplementary appendix A and the Shiny application can be found on GitHub (<https://github.com/iwadsworth/ElicitBiasPrior>). In practice, d_M and d_H could be chosen on the basis of adult dose-finding studies or, if the drug has already been licensed in adults, using World Health Organization lists of defined daily doses (World Health Organization Collaborating Centre for Drug Statistics Methodology, 2016). If an expert expresses the consistent opinion that the average response in younger children is similar to the fitted average in adolescents, this suggests that they believe that δ_A and δ_I are small, i.e. that the existing adult and adolescent data are highly relevant for informing our understanding of likely differences between adults and younger children. Note that the scheme proposed asks for opinions on the relevance of the existing data after an expert has seen how supportive they are of a complete extrapolation of efficacy data from adults to adolescents. To increase the credibility of beliefs that are elicited in this way, one could interview independent experts who are not directly involved with the drug development programme. Furthermore, the assumption that opinion on (δ_A, δ_I) is independent of opinion on other $E-R$ model parameters is a pragmatic one which ensures that elicitation questions have a direct interpretation and can be answered by non-statisticians.

By asking an expert for their best guesses at the average pharmacodynamic responses in children on placebo and a ‘high’ dose, we deduce fitted values of ν_A and ν_I . To find ν_A , one can subtract the fitted value of $\gamma_0 + \gamma_A$, obtained from the meta-analysis of existing data, from the expert’s best guess at the average response on placebo. Meanwhile, ν_I is obtained by subtracting the fitted value of $\gamma_C + \gamma_I$, also obtained from the meta-analysis, from the slope calculated by dividing the difference between the expert’s best guesses at the expected pharmacodynamic responses on a high dose and placebo by κd_H . An expert’s uncertainty about the average pharmacodynamic response in younger children on dose d^* , for $d^* \in (d_0, d_M, d_H)$, is established by asking questions to establish the fifth, 25th, 75th and 95th percentiles of their prior distribution for this quantity. Given values of ν_A and ν_I , we can then adapt the approach of Neuenschwander *et al.* (2008) to search over configurations of π_A^2 , π_I^2 and π_{AI} to find the triplet which defines a positive definite variance matrix and minimizes the absolute difference between percentiles of the fitted prior and the expert’s stated percentiles. To ensure positive definiteness, Π is represented in the optimization routine by using the Cholesky decomposition.

3.3. Example: application to antiepileptic drug development

The prior elicitation protocol and accompanying R Shiny application that are described in the on-line supplementary appendix A underwent several rounds of testing, with one author (IW) asking neurologists with experience of treating adult and/or paediatric epilepsy about dose-response curves for an antiepileptic drug. Testing included face-to-face pilot runs with eight

neurologists attending the International League Against Epilepsy British and Irish Chapters Meeting (Dublin, October 2016). The final version of the protocol was also piloted on three neurologists via web conference. The figures in supplementary appendix B are of the application tailored to the antiepileptic drug development application.

A few comments on the application of our elicitation scheme to the antiepileptic drug example are needed. The pharmacodynamic response, $Y = \log(Z + 110)$, is the log-transformed percentage change in seizure frequency from baseline. Since a log-transformed percentage change is difficult to give opinions on, we elicited beliefs on the percentage change in seizure frequency Z instead. It seems natural to think that, if we took an expert's best guess at the relationship between dose and $\mathbb{E}(Z)$ and then transformed it, we would obtain their best guess at the relationship between dose and $\mathbb{E}(Y)$. Therefore, the prior mode for $\mathbb{E}(Y)$ at a particular dose was obtained by transforming the prior mode of $\mathbb{E}(Z)$. The percentiles of an expert's prior distribution for $\mathbb{E}(Z)$ were similarly transformed to obtain percentiles of their prior for $\mathbb{E}(Y)$ (since the transformation was monotonic).

Using our elicitation procedure and Shiny application, bias priors E1 and E2 below were elicited from two epileptologists who were presented with simulated individual participant data on a licensed antiepileptic drug shown in Fig. 2. These were generated from the fitted models that were presented in Girgis *et al.* (2010). Assuming that $\kappa = 1$, the E – R –relationship is equivalent to the dose–response relationship:

(a) prior E1,

$$\begin{pmatrix} \delta_A \\ \delta_I \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.101^2 & 0 \\ 0 & 0.016^2 \end{pmatrix} \right\};$$

(b) prior E2,

$$\begin{pmatrix} \delta_A \\ \delta_I \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0.050 \\ 0.003 \end{pmatrix}, \begin{pmatrix} 0.101^2 & 0 \\ 0 & 0.016^2 \end{pmatrix} \right\}.$$

Prior E1 reflects the opinion that it is most likely that the average pharmacodynamic responses of adolescents and younger children are the same. Prior E2 is consistent with the belief that the E – R –curve in younger children lies slightly above that of adolescents (indicating a worse average response), so differences between E – R –curves in adults and younger children are larger than those between adults and adolescents. However, as can be seen from Fig. 2(b), both experts were uncertain about the dose–response curve in younger children given the existing data. The implications of this for extrapolation decisions are explored in Section 4.

4. Simulation study

Simulation scenarios were informed by applications to antiepileptic drug development for partial onset seizures (Girgis *et al.*, 2010; Nedelman *et al.*, 2007).

4.1. Epilepsy application extrapolation criterion

In all simulation scenarios, E – R –curves were said to be similar in two age groups if differences between median percentage changes from baseline in seizure frequency were less than 10%:

$$\mathcal{M}\{\exp(Y_T) - 110 | C_T = 0; A_T = 1\} - \mathcal{M}\{\exp(Y_T) - 110 | C_T = 0; A_T = 0\} \in (-10, 10) \quad (11)$$

and

$$\mathcal{M}\{\exp(Y_T) - 110 | C_T = C^*; A_T = 1\} - \mathcal{M}\{\exp(Y_T) - 110 | C_T = C^*; A_T = 0\} \in (-10, 10), \quad (12)$$

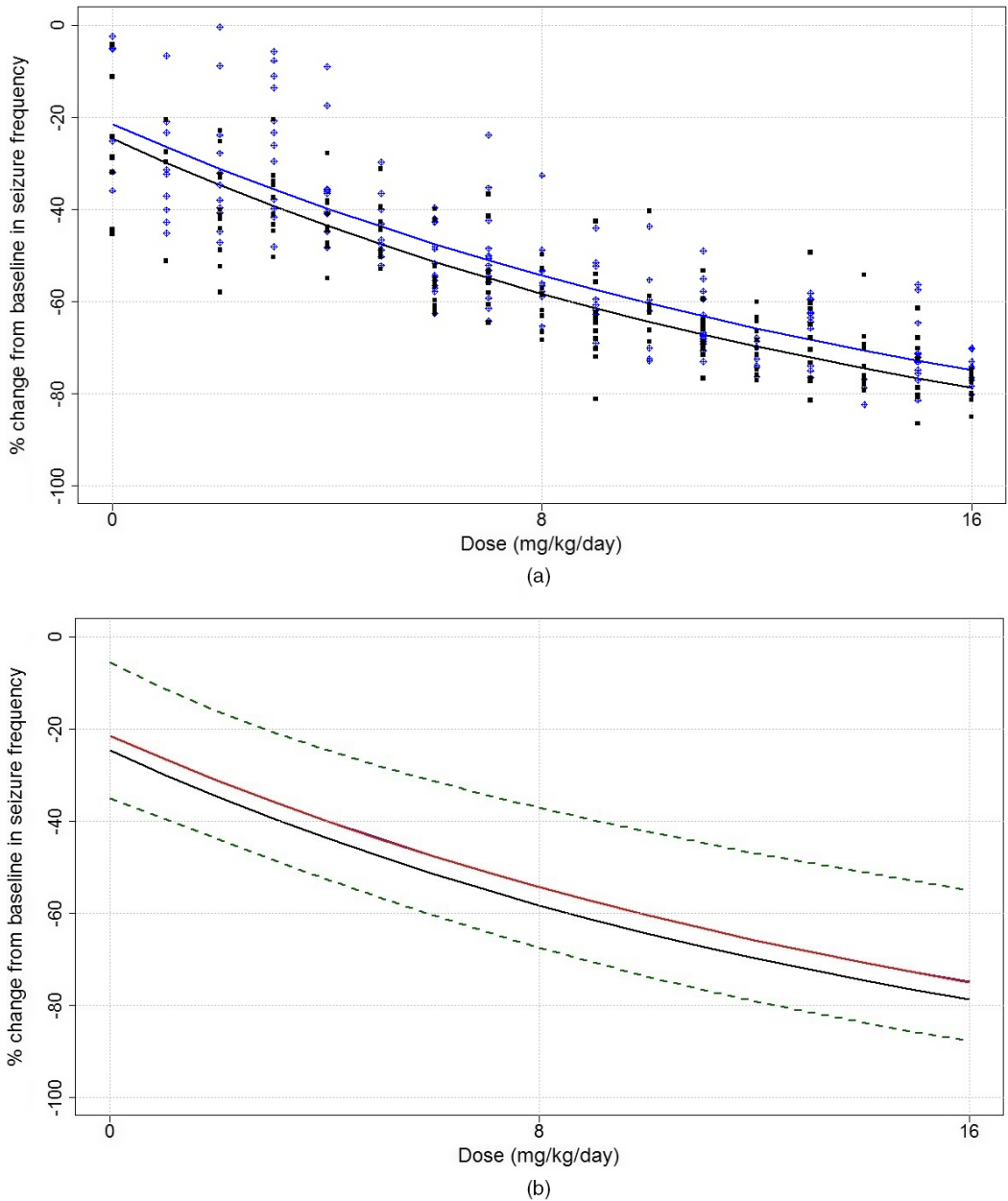


Fig. 2. (a) Simulated adult (■) and adolescent (◆) individual participant data and curves of best fit (—, adult; —, adolescent) and (b) elicited modal values of the dose–response relationship in younger children (—), with corresponding 90% credible interval (— —), used to derive prior E1: also plotted are adult (—) and adolescent (—) curves of best fit obtained from the simulated individual participant data

where \mathcal{M} represents the median and \mathcal{C}^* is the adult EC_{90} . Our choices for η_1 and η_2 were based on clinical feedback on acceptable differences in average responses. We wrote the similarity criteria in terms of the transformed pharmacodynamic end point to make it easier to elicit similarity bounds. We chose the median as our summary measure of response since, if Y follows a log-normal distribution with median m_Y , the median of $Z = \exp(Y) - 110$ is given by $m_Z = \exp(m_Y) - 110$, thus simplifying the mapping of properties from Z to Y .

4.2. Simulating existing E – R -data in adults and adolescents

We simulated the pharmacodynamic responses of adults and adolescents according to model (7). Setting the residual variance $\sigma^2 = 0.0243$ ensured that the transformed response Z lay within $\pm 10\%$ of its median given the patient’s age group and level of exposure with probability 0.95. We simulated age group indicators $A \sim \text{Bern}(0.15)$ so that on average 15% of existing trial participants were adolescents. This proportion appears reasonable on the basis of the studies that were cited in Girgis *et al.* (2010). Furthermore, we assigned 10% of patients in each study to placebo. For patients who were allocated to the drug, we sampled $\log(C_{\min})$ from a normal distribution with mean $\log(2.94)$ and variance 0.921, truncating samples above by $\log(17.27)$. In this way, we generated C_{\min} -values with quartiles and first and 99th percentiles similar to those reported by studies cited in Girgis *et al.* (2010) where C_{\min} -values ranged between 0.19 and $17.27 \mu\text{g ml}^{-1}$.

For each existing study, study-specific parameters of E – R -model (7) were generated by sampling $\gamma_{0j} \sim N(\gamma_0, \sigma_0^2)$, $\gamma_{Cj} \sim N(\gamma_C, \sigma_C^2)$ and

$$\begin{pmatrix} \gamma_{Aj} \\ \gamma_{Ij} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \gamma_A \\ \gamma_I \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_I^2 \end{pmatrix} \right\},$$

setting $\gamma_0 = 4.4469$ and $\gamma_C = -0.0627$, which are the maximum likelihood estimates of these parameters taken from Girgis *et al.* (2010). Let ϵ_P and ϵ_C represent the difference between $\mathcal{M}(Z|A = 1, C)$ and $\mathcal{M}(Z|A = 0, C)$ when $C = 0$ and $C = C^*$ respectively. We chose values for γ_A and γ_I such that ϵ_P and ϵ_C , when evaluated under these average parameters, spanned a realistic range of differences. We considered pairs $(\epsilon_P, \epsilon_C) \in \{(0, 0), (5, 5), (10, 10), (20, 20), (5, 10), (5, 20)\}$ which correspond to the six pairs of (γ_A, γ_I) labelled in Table 1 as E – R -models S1–S6.

Table 1. Population means of the effects of age, γ_A , and the interaction between age and exposure, γ_I , for adults and adolescents in the six E – R simulation models, with the interpretation of each model†

Model	γ_A	γ_I	Model interpretation
S1	0	0	Population median pharmacodynamic response identical in adults and adolescents
S2	0.057	0.006	Small differences between population median pharmacodynamic responses satisfy criteria (11)–(12)
S3	0.111	0.010	Moderate differences satisfy criteria (11)–(12)
S4	0.211	0.018	Large differences do not satisfy criteria (11)–(12)
S5	0.057	0.014	Small differences between population median pharmacodynamic responses on placebo; moderate differences at EC_{90} satisfy criteria (11)–(12)
S6	0.057	0.027	Small differences between population median pharmacodynamic responses on placebo; large differences at EC_{90} do not satisfy criterion (12)

†The population median pharmacodynamic response refers to $\exp(m_Y) - 110$, where $m_Y = \mathcal{M}(Y)$ is calculated by setting E – R model parameters equal to their population means.

The variances of study-specific E – R -parameters were chosen to characterize low, moderate, high and very high levels of between-trial heterogeneity. We chose σ_0^2 such that a study-specific value of $\mathcal{M}(Z|A=0, C=0)$ lay within $\pm 10\%$ of the median of Z calculated by setting the E – R -model parameters equal to their population means with probability 0.6 (very high heterogeneity), 0.7 (high), 0.8 (moderate) or 0.95 (low). Fixing σ_0^2 , σ_C^2 was then set to ensure that the study-specific value of $\mathcal{M}(Z|A=0, C=EC_{90})$ lay within $\pm 10\%$ of the median of Z calculated by setting the E – R -model parameters equal to their population means with the same probability. We chose σ_A^2 and σ_1^2 to fix the probability that an individual existing trial will be consistent with an assumption of similar E – R -curves in adults and adolescents according to criteria (11)–(12). Specifically, we chose σ_A^2 such that, with probability 0.6, 0.7, 0.8 or 0.95, the true difference in a study between $\mathcal{M}(Z|A=0, C=0)$ and $\mathcal{M}(Z|A=1, C=0)$ lay within $\pm 10\%$. For a particular choice of σ_A^2 , we then fixed σ_1^2 such that, with probability 0.6, 0.7, 0.8 or 0.95, the true difference between study-specific values $\mathcal{M}(Z|A=0, C=EC_{90})$ and $\mathcal{M}(Z|A=1, C=EC_{90})$ lay within $\pm 10\%$. Different configurations of the heterogeneity parameters are listed in Table 2.

Simulation scenarios considered different numbers of existing trials ($H=2, 3, 4, 5, 10, 20$) and numbers of subjects per trial ($N=30, 170$). Numbers of existing trials were chosen to explore a plausible range: Davey *et al.* (2011) reported that 75% of the 22453 meta-analyses that were listed in the Cochrane database of systematic reviews at 2011 were based on five or fewer studies, and 1% were based on 28 or more. Values of N were informed by four industry-sponsored trials of an antiepileptic drug, the average sample size of which was 168 patients.

4.3. Meta-analysis of simulated existing E – R -studies

For each of the 288 simulation scenarios, we simulated 1000 sets of trials and fitted the Bayesian meta-analytic model of Section 2.3 to each data set. All simulations were performed in R (R Development Core Team, 2015) fitting the meta-analytic model by calling OpenBUGS version 3.2.3 (Lunn *et al.*, 2009) using the R2OpenBUGS package (Sturtz *et al.*, 2005). We fitted the Bayesian model by running three chains using a thinning rate of 5, running the chain for 30000 iterations including a burn-in of 10000 iterations. The coda package (Plummer *et al.*, 2006) was then used to extract posterior samples from the OpenBUGS output.

The meta-analytic model was fitted stipulating the priors that are given in Table 3. Hyperparameters for ξ_1 and ξ_2 , defining the variability of the random-effects distribution (8), were chosen so that $\mathbb{E}(\xi_1)$ and $\mathbb{E}(\xi_2)$ were equal to our choices for the moderate between-trial standard deviation for γ_A and γ_1 given in Table 2. Then 95% of the probability mass for the ξ_1 -prior was between (0.012, 0.250), and 95% of the probability mass for the ξ_2 -prior was between (0.001, 0.050). Therefore, low weight was given to very low and high between-trial variances. We

Table 2. Standard deviations for the intercept σ_0 and effects of exposure, σ_C , age, σ_A , and the interaction between age and exposure, σ_1 , chosen to reflect low, moderate, high and very high levels of between-trial heterogeneity

Level of heterogeneity	σ_0	σ_C	σ_A	σ_1
Low	0.057	0.008	0.059	0.009
Moderate	0.086	0.012	0.091	0.014
High	0.107	0.015	0.113	0.018
Very high	0.132	0.019	0.140	0.023

Table 3. Prior distributions placed on parameters of model (7) and random-effects distribution (8)

Parameter	Prior distribution
$\gamma_{01}, \dots, \gamma_{0H}$	$N(0, 100)$
$\gamma_{C1}, \dots, \gamma_{CH}$	$N(0, 100)$
ξ_1	$\text{gamma}(2.097, 23.003)$
ξ_2	$\text{gamma}(1.118, 78.149)$
λ_0	$t(0, 0.03, \text{df}=3)$
λ_1	$N(0, 1)$

repeated some simulations placing $\text{gamma}(1.5, 3)$ priors on ξ_1 and ξ_2 , which resulted in a slight reduction in the probability of extrapolation.

5. Results

5.1. Meta-analysis of existing data

As can be seen from on-line supplementary Tables ST9–ST16, the Bayesian multivariate meta-analysis of the existing adult and adolescent data produces accurate estimates of γ_A and γ_I , with low bias, empirical standard deviations and mean-squared error in most scenarios. In all cases, the accuracy increases with the sample size per existing study. Empirical standard deviations are highest under the highest level of between-trial heterogeneity, but the bias remains small. The intercept and effect of exposure are also estimated with small bias and high precision (the results are not presented).

5.2. Effective sample sizes of the approximate joint prior for parameters representing differences between adults and younger children

On-line supplementary appendix C explores how the average effective sample sizes of the bivariate normal mixture approximation to the joint prior for (β_A, β_I) is influenced by an expert’s uncertainty about the bias parameters. Under bias prior E1, information from the existing adult and adolescent data is heavily downweighted; for example, if each existing study enrolled 170 patients, assuming low between-trial heterogeneity, the effective sample size of the prior for (β_A, β_I) would be 24. The effective sample size of the prior for (β_A, β_I) increases as prior uncertainty about the external biases decreases. For more details, the reader is referred to supplementary appendix C.

5.3. Prior probability that E–R-curves are similar in adults and younger children

First we look at how the prior probability of the extrapolation criteria (11)–(12) (referred to as p_E) varies with the true E–R-relationship in adults and adolescents. On-line supplementary Tables ST17–ST20 present the means and empirical standard deviations of p_E for a range of scenarios. Figs 3(d)–3(f) illustrate E–R-relationships in adults, adolescents and younger children under simulation models S1, S3 and S4 and bias prior E1. Figs 3(a)–3(c) illustrate how p_E changes as differences between adult and adolescent E–R-relationships increase from none (model S1), to moderate (model 3) to large (model 4) under bias prior E1. A general trend is that larger values of p_E are recorded in scenarios where the true E–R-curves in adults and adolescents are more closely aligned. Under bias prior E1 and models S1 and S3, when differences between

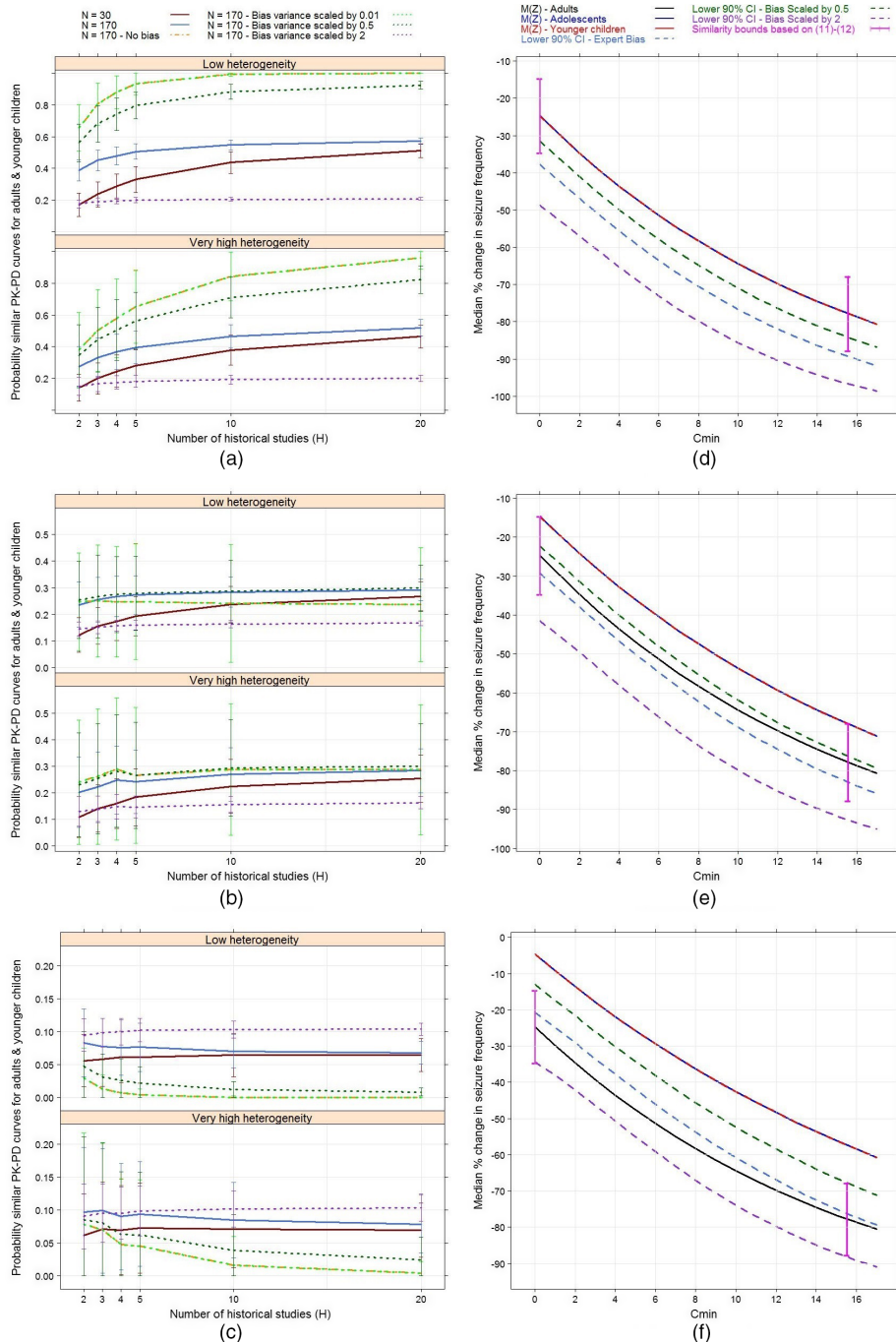


Fig. 3. (a)–(c) Average prior extrapolation probabilities under bias prior E1 and simulation scenarios S1, S3 and S4 respectively (\bar{p} , ± 1 empirical standard deviation of the observed probabilities) and (d)–(f) median responses in adults, adolescents and younger children under simulation models S1, S3 and S4 respectively: also plotted are the lower bounds of 90% credible intervals for the median response in younger children, consistent with bias prior E1 when its variance matrix is unscaled, or scaled by a factor of 0.5 or 2; credible intervals are calculated conditioning on true values of adult and adolescent $E-R$ -parameters; bars at the placebo and EC_{90} represent similarity bounds given criteria (11) and (12)

adults and adolescents are sufficiently small to satisfy criteria (11) and (12), p_E increases as N increases, although differences diminish with H .

Under model S1 and bias prior E1, p_E reaches a maximum of 0.572 when the between-trial heterogeneity is low and data are available from $H = 20$ existing studies, each having recruited $N = 170$ subjects. Curves representing cases when the bias prior variance matrix is scaled by a factor of 0.5, 0.01 or 0 show that, if uncertainty about external biases were to be significantly reduced, p_E would increase. If the prior variance matrix is scaled by a factor of 0, we take this to mean that the expert knows that $\delta_A = \nu_A$ and $\delta_I = \nu_I$; therefore, their joint prior places probability mass 1 on the configuration $(\delta_A, \delta_I) = (\nu_A, \nu_I)$ and assigns zero probability to all other pairs. For bias prior E1, a scale factor of 0 would reflect the opinion that we are certain that differences between adult and adolescent $E-R$ -curves reflect differences between curves for adults and younger children. There is a question of whether it is plausible that an expert would be sufficiently confident in their beliefs for us to attain a high value of p_E . Suppose that $p_E = 0.8$ would be sufficient to support a complete extrapolation strategy. From Fig. 3(a), we see that, under model S1 with low between-trial heterogeneity and 170 subjects per trial, if we scale the bias prior variance matrix by 0.5, p_E reaches 0.8 when $H > 5$. From Fig. 3(d), we can see what this scaling factor would correspond to in terms of the level of confidence that an expert must have in the location of the $E-R$ -curve in younger children. We speculate that experts could have this level of confidence in practice. With enough existing data and strong, but still feasible, expert opinion, a high prior extrapolation probability is plausible.

Prior probabilities of extrapolation under models S2, S5 and S6 are provided in the on-line supplementary materials as Figs SF9–SF11. Similar patterns are seen in the results that are generated under models S2 and S5 to those under model S1, although values of p_E tend to be lower overall, reflecting the larger differences between adult and adolescent $E-R$ -curves. A similar comment applies to results generated under models S4 and S6.

We have repeated our investigations by using bias prior E2. Comparing results generated under the two priors, we see that the prior probability of extrapolation under prior E2 is lower in all scenarios, demonstrating that it is not only an expert's uncertainty about external biases which influences the probability of extrapolation but also the expert's opinion on the direction of differences between $E-R$ -curves in adolescents and younger children.

6. Discussion

This paper proposes a quantitative framework for using existing pharmacological data to inform our understanding of likely differences between $E-R$ -relationships in adults and younger children. The prior probability of acceptably small differences between these relationships is used to inform a decision of whether to perform a complete or partial extrapolation of adult efficacy data to younger children. Currently, we propose that prior extrapolation probabilities in excess of 0.8 or 0.9 would support a decision to adopt a complete extrapolation strategy, although further work will explore whether the choice of this cut-off can be refined and formalized through the use of a decision theoretic argument. Such an approach would consider various risks and costs, including the risk to children of incorrectly adopting a complete extrapolation strategy and the costs to patients and the sponsor of failing to perform a complete extrapolation when this is appropriate. When it is unclear whether a complete extrapolation strategy should be adopted or not, one could use an expected value of information analysis (Briggs *et al.*, 2006; Heath *et al.*, 2017) to quantify the value, in terms of improved decision making, of collecting varying numbers of additional $E-R$ -data in younger children given the risks that were outlined above.

When performing the bias-adjusted meta-analysis on which the prior probability of extrapolation is based, it is essential that the studies included should have been identified through a process of systematic review according to a prespecified protocol (Higgins and Green, 2011; Khan *et al.*, 2011). We suggest that the eligibility criteria for the systematic review should include trials that recruit patients with the indication of interest, assess the same drug and collect follow-up data that allow the chosen outcomes to be analysed. To inform expert opinion, during the elicitation meeting we could broaden out to present data on related drugs or indications.

Our current approach assumes that E – R -relationships can be captured by models which represent age as a categorical variable, i.e. that assume that there are no important differences within an age group. Although this assumption will never hold exactly, we do expect it to hold approximately for suitably defined age groups. If important differences were expected to occur within an age group, then a more suitable approach would be to consider each homogeneous age group in turn and to select an extrapolation strategy for each by application of the methods that were described in Sections 2 and 3. Although the motivating example for this work has been extrapolating across age groups, a similar framework could be used to inform the extrapolation of efficacy data across ethnic groups or geographic regions, where subgroups in this setting are naturally discrete.

In this paper, we have considered the case of linear E – R -models, though generalized linear models would be accommodated in our framework with appropriate adjustments to the elicitation protocol. However, over the therapeutic window of interest, the E – R -relationship is likely to be approximately linear in many cases, even if the complete E – R -relationship follows an Emax model (Macdougall, 2006): a non-linear model which is often used to model the relationship between exposure and response. If this is not so, extending to non-linear E – R -models could be possible, though one would need to consider

- (a) how to parameterize the more complex E – R -models for adults, adolescents and younger children,
- (b) how to represent differences between the various E – R -relationships and to define decision criteria governing extrapolation decisions and
- (c) how one would devise a scheme to elicit opinion on biases affecting parameters governing the similarity of E – R -relationships.

Additionally, our prior extrapolation probability would probably need to consider a moderate exposure level, such as EC_{50} , along with a placebo and a higher exposure.

Acknowledgements

We are grateful to acknowledge funding from the Medical Research Council (grant MR/M013510/1). Dr Wadsworth's contribution to the paper was made while he was an employee of Lancaster University. Dr Hampson's contribution to the manuscript was made while she was an employee of Lancaster University and AstraZeneca. 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 that are 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 also thank the experts who took part with trialling the prior elicitation scheme (Drs Helen Cross, Yvonne Hart, Nikolas Hitiris, Stephen Howell, Charlotte Lawthom, Rajiv Mohanraj, Graham Powell, Simon Shorvon, David Smith, William Whitehouse and Sameer Zuberi), and Dr Björn Bornkamp and Dr Linh Van for helpful discussions.

References

- Briggs, A. H., Claxton, K. and Sculpher, M. J. (2006) *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press.
- Crippa, A., Discacciati, A., Bottai, M., Spiegelman, D. and Orsini, N. (2018) One-stage dose-response meta-analysis for aggregated data. *Statist. Meth. Med. Res.*, **28**, 1579–1596.
- Davey, J., Turner, R. M., Clarke, M. J. and Higgins, J. P. (2011) Characteristics of meta-analyses and their component studies in the Cochrane database of systematic reviews: a cross-sectional, descriptive analysis. *BMC Med. Res. Methodol.*, **11**, 1–11.
- Dunne, J., Rodriguez, W. J., Murphy, M. D., Barker, P., Sproule, S., Beasley, B. N., Burckart, G. J., Filie, J. D., Lewis, L. L., Sachs, H. C., Sheridan, P. H., Starke, P. and Yao, L. P. (2011) Extrapolation of adult data and other data in pediatric drug-development programs. *Pediatrics*, **128**, 1242–1249.
- European Medicines Agency (2010) Conclusions of the paediatric epilepsy experts group meeting, held in London 1 September 2009. *Report EMA/153272/2010*. European Medicines Agency, Amsterdam. (Available from http://www.ema.europa.eu/docs/en_GB/document_library/Other/2010/05/WC500090218.pdf.)
- European Medicines Agency (2013) Concept paper on extrapolation of efficacy and safety in medicine development (Final). *Report EMA/129698/2012*. European Medicines Agency, Amsterdam. (Available from http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/04/WC500142358.pdf.)
- European Medicines Agency (2017) Reflection paper on the use of extrapolation in the development of medicines for paediatrics (Draft). *Report EMA/199678/2016*. European Medicines Agency, Amsterdam. (Available from http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/10/WC500236640.pdf.)
- European Union (2006) Regulation (EC) No 1901/2006 of the European Parliament and of the Council of 12 December 2006 on medicinal products for paediatric use and amending Regulation (EEC) No 1768/92, Directive 2001/20/EC, Directive 2001/83/EC and Regulation (EC) No 726/2004. *Official Journal L 378*, 27/12/2006, pp. 1–19.
- FitzGerald, J. M., Bleecker, E. R., Nair, P., Korn, S., Ohta, K., Lommatzsch, M., Ferguson, G. T., Busse, W. W., Gilmartin, G., Werkström, V., Aurivillius, M., Goldman, M. and the CALIMA Study Investigators (2016) Benralizumab, an anti-interleukin-5 receptor α monoclonal antibody, as add-on treatment for patients with severe, uncontrolled, eosinophilic asthma (CALIMA): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet*, **388**, 2128–2141.
- Food and Drug Administration (2003) Guidance for industry: exposure-response relationships—study design, data analysis, and regulatory applications. Food and Drug Administration, Rockville. (Available from <http://www.gmp-compliance.org/guidemgr/files/EXPOSURE.pdf>.)
- Food and Drug Administration (2018) Considerations for the inclusion of adolescent patients in adult oncology clinical trials guidance for industry. Food and Drug Administration, Rockville. (Available from <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM609513.pdf>.)
- French, J. A., Kanner, A. M., Bautista, J., Abou-Khalil, B., Browne, T., Harden, C. L., Theodore, W. H., Bazil, C., Stern, J., Schachter, S. C., Bergen, D., Hirtz, D., Montouris, G. D., Nespeca, M., Gidal, B., Marks, Jr, W. J., Turk, W. R., Fischer, J. H., Bourgeois, B., Wilner, A., Faught, Jr, R. E., Sachdeo, R. C., Beydoun, A., Glauser, T. A., Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology, Quality Standards Subcommittee of the American Academy of Neurology and American Epilepsy Society (2004) Efficacy and tolerability of the new antiepileptic drugs, II: treatment of refractory epilepsy: report of the TTA and QSS subcommittees of the American Academy of Neurology and the American Epilepsy Society. *Epilepsia*, **45**, 410–423.
- French, J. A., Krauss, G. L., Biton, V., Squillacote, D., Yang, H., Laurenza, A., Kumar, D., Rogawski, M. A., Campanille, V. and Floridia, J. (2012) Adjunctive perampanel for refractory partial-onset seizures randomized phase III study 304. *Neurology*, **79**, 589–596.
- Galwey, N. (2017) Supplementation of a clinical trial by historical control data: is the prospect of dynamic borrowing an illusion? *Statist. Med.*, **36**, 899–916.
- Gamalo-Siebers, M., Savic, J., Basu, C., Zhao, X., Gopalakrishnan, M., Gao, A., Song, G., Baygani, S., Thompson, L., Xia, H. A., Price, K., Tiwari, R. and Carlin, B. P. (2017) Statistical modeling for Bayesian extrapolation of adult clinical trial information in pediatric drug evaluation. *Pharm. Statist.*, **16**, 232–249.
- Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models. *Bayes Anal.*, **1**, 515–534.
- Girgis, I. G., Nandy, P., Nye, J. S., Ford, L., Mohanty, S., Wang, S., Ochalski, S., Eerdekens, M. and Cox, E. (2010) Pharmacokinetic-pharmacodynamic assessment of topiramate dosing regimens for children with epilepsy 2 to < 10 years of age. *Epilepsia*, **51**, 1954–1962.
- Grün, B. and Leisch, F. (2007) Fitting finite mixtures of generalized linear regressions in R. *Computat. Statist. Data Anal.*, **51**, 5247–5252.
- Grün, B. and Leisch, F. (2008) FlexMix version 2: finite mixtures with concomitant variables and varying and constant parameters. *J. Statist. Softw.*, **28**, 1–35.

- Hampson, L. V., Whitehead, J., Eleftheriou, D. and Brogan, P. (2014) Bayesian methods for the design and interpretation of clinical trials in very rare diseases. *Statist. Med.*, **33**, 4186–4201.
- Heath, A., Manolopoulou, I. and Baio, G. (2017) A review of methods for analysis of the expected value of information. *Med. Decsn Makng*, **37**, 747–758.
- Higgins, J. P. T. and Green, S. (2011) *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. Cochrane Collaboration. (Available from www.handbook.cochrane.org.)
- Hlavin, G., Koenig, F., Male, C., Posch, M. and Bauer, P. (2016) Evidence, eminence and extrapolation. *Statist. Med.*, **35**, 2117–2132.
- Hobbs, B. P., Carlin, B. P., Mandrekar, S. J. and Sargent, D. J. (2011) Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, **67**, 1047–1056.
- Ibrahim, J. G. and Chen, M.-H. (2000) Power prior distributions for regression models. *Statist. Sci.*, **15**, 46–60.
- International Conference on Harmonisation (2001) E11: note for guidance on clinical investigation of medicinal products in the paediatric population (CPMP/ICH/2711/99). European Medicines Agency, Amsterdam. (Available from http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002926.pdf.)
- Kearns, G. L., Abdel-Rahman, S. M., Alander, S. W., Blowey, D. L., Leeder, J. S. and Kauffman, R. E. (2003) Developmental pharmacology—drug disposition, action, and therapy in infants and children. *New Engl. J. Med.*, **349**, 1157–1167.
- Khan, K., Kunz, R., Kleijnen, J. and Antes, G. (2011) *Systematic Reviews to Support Evidence-based Medicine*. Boca Raton: CRC Press.
- Kullback, S. and Leibler, R. A. (1951) On information and sufficiency. *Ann. Math. Statist.*, **22**, 79–86.
- Leisch, F. (2004) FlexMix: a general framework for finite mixture models and latent class regression in R. *J. Statist. Softw.*, **11**, 1–18.
- Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009) The BUGS project: evolution, critique and future directions. *Statist. Med.*, **28**, 3049–3067.
- Macdougall, J. (2006) Analysis of dose–response studies—Emax model. In *Dose Finding in Drug Development* (ed. N. Ting). New York: Springer.
- Marshall, J. D. and Kearns, G. L. (1999) Developmental pharmacodynamics of cyclosporine. *Clin. Pharmacol. Therapeut.*, **66**, 66–75.
- Marson, A. G., Al-Kharusi, A. M., Alwaidh, M., Appleton, R., Baker, G. A., Chadwick, D. W., Cramp, C., Cockerell, O. C., Cooper, P. N., Doughty, J., Eaton, B., Gamble, C., Goulding, P. J., Howell, S. J., Hughes, A., Jackson, M., Jacoby, A., Kellett, M., Lawson, G. R., Leach, J. P., Nicolaides, P., Roberts, R., Shackley, P., Shen, J., Smith, D. F., Smith, P. E., Smith, C. T., Vanoli, A., Williamson, P. R. and the SANAD Study Group (2007a) The SANAD study of effectiveness of carbamazepine, gabapentin, lamotrigine, oxcarbazepine, or topiramate for treatment of partial epilepsy: an unblinded randomised controlled trial. *Lancet*, **369**, 1000–1015.
- Marson, A. G., Al-Kharusi, A. M., Alwaidh, M., Appleton, R., Baker, G. A., Chadwick, D. W., Cramp, C., Cockerell, O. C., Cooper, P. N., Doughty, J., Eaton, B., Gamble, C., Goulding, P. J., Howell, S. J., Hughes, A., Jackson, M., Jacoby, A., Kellett, M., Lawson, G. R., Leach, J. P., Nicolaides, P., Roberts, R., Shackley, P., Shen, J., Smith, D. F., Smith, P. E., Smith, C. T., Vanoli, A., Williamson, P. R. and the SANAD Study Group (2007b) The SANAD study of effectiveness of valproate, lamotrigine, or topiramate for generalised and unclassifiable epilepsy: an unblinded randomised controlled trial. *Lancet*, **369**, 1016–1026.
- Medical Research Council Biostatistics Unit (2017) The BUGS project FAQs. Medical Research Council Biostatistics Unit, Cambridge. (Available from <http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-faqs/>.)
- Nedelman, J. R., Rubin, D. B. and Sheiner, L. B. (2007) Diagnostics for confounding in PK/PD models for oxcarbazepine. *Statist. Med.*, **26**, 290–308.
- Neuenschwander, B., Branson, M. and Gsponer, T. (2008) Critical aspects of the Bayesian approach to phase I cancer trials. *Statist. Med.*, **27**, 2420–2439.
- Neuenschwander, B., Capkun-Niggli, G., Branson, M. and Spiegelhalter, D. J. (2010) Summarizing historical information on controls in clinical trials. *Clin. Trials*, **7**, 5–18.
- O’Byrne, P. M., FitzGerald, J. M., Zhong, N., Bateman, E., Barnes, P. J., Keen, C., Almqvist, G., Pemberton, K., Jorup, C., Ivanov, S. and Reddel, H. K. (2017) The SYGMA programme of phase III trials to evaluate the efficacy and safety of budesonide/formoterol given ‘as needed’ in mild asthma: study protocols for two randomised controlled trials. *Trials*, **18**, 12.
- Pediatric News (2016) FDA conducts analysis to assess acceptability of extrapolation of antiepileptic drug (AED) effectiveness in adults to children four years of age and older with partial onset seizures (POS). *J. Pediatr. Pharmacol. Ther.*, **21**, 98.
- Petit, C., Samson, A., Morita, S., Ursino, M., Guedj, J., Jullien, V., Comets, E. and Zohar, S. (2018) Unified approach for extrapolation and bridging of adult information in early-phase dose-finding paediatric studies. *Statist. Meth. Med. Res.*, **27**, 1860–1877.
- Plummer, M., Best, N., Cowles, K. and Vines, K. (2006) CODA: convergence diagnosis and output analysis for MCMC. *R News*, **6**, 7–11.
- R Development Core Team (2015) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

- Rietbergen, C., Klugkist, I., Janssen, K. J., Moons, K. G. and Hoijtink, H. J. (2011) Incorporation of historical data in the analysis of randomized therapeutic trials. *Contemp. Clin. Trials*, **32**, 848–855.
- Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D. and Neuenschwander, B. (2014) Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, **70**, 1023–1032.
- Stephenson, T. (2005) How children's responses to drugs differ from adults. *Br. J. Clin. Pharmacol.*, **59**, 670–673.
- Sturtz, S., Ligges, U. and Gelman, A. (2005) R2WinBUGS: a package for running WinBUGS from R. *J. Statist. Softw.*, **12**, 1–16.
- Sun, H., Temeck, J. W., Chambers, W., Perkins, G., Bonnel, R. and Murphy, D. (2017) Extrapolation of efficacy in pediatric drug development and evidence-based medicine: progress and lessons learned. *Therapeut. Innov. Regltry Sci.*, 1–7.
- Takahashi, H., Ishikawa, S., Nomoto, S., Nishigaki, Y., Ando, F., Kashima, T., Kimura, S., Kanamori, M. and Echizen, H. (2000) Developmental changes in pharmacokinetics and pharmacodynamics of warfarin enantiomers in Japanese children. *Clin. Pharmacol. Therapeut.*, **68**, 541–555.
- Tan, S.-B., Dear, K. B., Bruzzi, P. and Machin, D. (2003) Strategy for randomised clinical trials in rare cancers. *Br. Med. J.*, **327**, 47–49.
- Turner, R. M., Spiegelhalter, D. J., Smith, G. C. S. and Thompson, S. G. (2009) Bias modelling in evidence synthesis. *J. R. Statist. Soc. A*, **172**, 21–47.
- Wadsworth, I., Hampson, L. V. and Jaki, T. (2016a) Extrapolation of efficacy and other data to support the development of new medicines for children: a systematic review of methods. *Statist. Meth. Med. Res.*, **27**, 398–413.
- Wadsworth, I., Jaki, T., Sills, G. J., Appleton, R., Cross, J. H., Marson, A. G., Martland, T., McLellan, A., Smith, P. E., Pellock, J. M. and Hampson, L. V. (2016b) Clinical drug development in epilepsy revisited: a proposal for a new paradigm streamlined using extrapolation. *CNS Drugs*, **30**, 1011–1017.
- Weber, K., Hemmings, R. and Koch, A. (2018) How to use prior knowledge and still give new data a chance? *Pharmaceut. Statist.*, **17**, 329–341.
- Welton, N. J., Ades, A. E., Carlin, J. B., Altman, D. G. and Sterne, J. A. C. (2009) Models for potentially biased evidence in meta-analysis using empirically based priors. *J. R. Statist. Soc. A*, **172**, 119–136.
- Willan, A. R. and Pinto, E. M. (2005) The value of information and optimal clinical trial design. *Statist. Med.*, **24**, 1791–1806.
- Wilson, E. C. (2015) A practical guide to value of information analysis. *PharmacoEconomics*, **33**, 105–121.
- World Health Organization Collaborating Centre for Drug Statistics Methodology (2016) Defined daily dosing—definition and general considerations. World Health Organization, Geneva. (Available from https://www.whocc.no/ddd/definition_and_general_considera/.)
- Zheng, H., Hampson, L. V. and Wandel, S. (2019) A robust Bayesian meta-analytic approach to incorporate animal data into phase I oncology trials. *Statist. Meth. Med. Res.*, to be published, doi 10.1177/0962280218820040.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Web-based supporting materials for "A quantitative framework to inform extrapolation decisions in children".'