Statistical Modeling for Bayesian Extrapolation of Adult Clinical Trial Information in Pediatric Drug Evaluation

Margaret Gamalo-Siebers,[†] Jasmina Savic, Cynthia Basu, Xin Zhao, Mathangi Gopalakrishnan, Aijun Gao, Guochen Song, Simin Baygani, Laura Thompson, H. Amy Xia, Karen Price, Ram Tiwari, and Bradley P. Carlin
for the DIA Bayesian Statistics Working Group

August 28, 2016

Abstract

Children represent a large under-served population of "therapeutic orphans", as an estimated 80% of children are treated off-label. However, pediatric drug development often faces substantial challenges, including economic, logistical, technical, and ethical barriers, among others. Among many efforts trying to remove these barriers, increased recent attention has been paid to *extrapolation*; that is, the leveraging of available data from adults or older age groups to draw conclusions for the pediatric population. The Bayesian statistical paradigm is natural in this setting, as it permits the combining (or "borrrowing") of information across disparate sources, such as the adult and pediatric data. In this paper, authored by the pediatric subteam of the Drug Information Association (DIA) Bayesian Scientific Working Group and Adaptive Design Working Group, we develop, illustrate, and provide suggestions on Bayesian statistical methods that could be used to design improved pediatric development programs that use all available information in the most efficient manner. A variety

[†]Corresponding Author: Margaret Gamalo-Siebers, PhD, Principal Research Scientist, Advanced Analytics, Eli Lilly & Co, Lilly Corporate Center, Indianapolis, IN 46285; Email: gamalo_margaret@lilly.com. Jasmina Savic, MS, RAC is a Regulatory Affairs Consultant at JS Regulatory, Mariahilfstr. 16, 52062 Aachen, Germany. Cynthia Basu, MS is a graduate student at the Division of Biostatistics, University of Minnesota, 420 Delaware St. SE. Minneapolis, MN 55455. Xin Zhao, PhD is Director, Biostatistics, Johnson & Johnson, 32 Oyster Point Boulevard, South San Francisco, CA 94080. Mathangi Gopalakrishnan, PhD is Research Assistant Professor, Center for Translational Medicine, University of Maryland, 620 W Lexington St, Baltimore, MD 21201. Aijun Gao, PhD is Associate Director, Biostatistics, Chiltern International Ltd., 1016 West Ninth Avenue, King of Prussia, PA 19406. Guochen Song, PhD is Associate Director, Biogen, 225 Binney Street Cambridge, MA 02142. Simin Baygani, MS is Reseach Scientist, Global Statistical Science, Eli Lilly & Co, Lilly Corporate Center, Indianapolis, IN 46285. Laura Thompson, PhD is a Mathematical Statistician, Office of Surveillance and Biometrics, Center for Devices and Radiological Health, Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD 20993. H. Amy Xia, PhD is Executive Director, Biostatistics, Amgen, One Amgen Center Drive Thousand Oaks, CA 91320. Karen Price, PhD is Research Advisor, Advanced Analytics, Eli Lilly & Co, Lilly Corporate Center, Indianapolis, IN 46285. Ram Tiwari, PhD is Associate Director, Office of Biostatistics, Center for Drug Evaluation and Research, Food and Drug Administration, 10903 New Hampshire Ave, Silver Spring, MD 20993. Bradley P. Carlin, PhD is Chair, Division of Biostatistics, University of Minnesota, 420 Delaware St. SE. Minneapolis, MN 55455. This article reflects the views of the authors and should not be construed to represent FDA's views or policies or the views or policies of the author's employers.

of relevant Bayesian approaches are described, several of which are illustrated through two case studies: extrapolating adult efficacy data to expand the labeling for Remicade® to include pediatric ulcerative colitis, and extrapolating adult exposure-response information for anti-epileptic drugs to pediatrics.

Keywords: Commensurate prior; exchangeability; extrapolation; effective sample size; hierarchical model; model fit; power prior.

1 Introduction

1.1 Pediatric Drug Development

The Scientific Working Groups (SWG) is a community within the Drug Information Association (DIA) that focuses on developing innovative methods and tools and on promoting the dissemination of such methods and best practices. One area that has been increasingly recognized within the group is the need for new and improved statistical methods useful in the development of medicinal products for pediatric populations. As such, an effort was formally organized within the Adaptive Design SWG, the Bayesian SWG, and the pediatric community in 2014 and 2015 to develop recommendations for optimal use of statistical methods in development programs of new medicinal products in pediatrics.

This paper resulted through the voluntary contributions of the pharmaceutical company, Contract Research Organization (CRO), academic, and federal agency SWG statisticians who provided their perspectives on improved pediatric development program methods that would be more efficient while maintaining statistical integrity. We hope to offer clearer solutions to commonly occurring obstacles that have prevented industry confidence in this area.

Pediatric drug development faces substantial hurdles, including economic, logistical, technical, and ethical barriers, among others [1]. While the economic challenges have been addressed with the enactment of pediatric regulations, e.g., Best Pharmaceuticals for Children Act (BPCA) [2] and the Pediatric Research Equity Act (PREA) [3], which aim to increase knowledge on the use of pediatric drugs through a system of obligations and incentives, other barriers remain. On the technical side, for example, even when data on drugs successfully tested on adults are available, we may still lack information on the dosing, safety, and/or efficacy of these drugs in children. Previous research demonstrates that children may not respond to medications in the same way as adults (see e.g. [4] and [5]). Differences between children and adults include ways in which medicines are adsorbed, distributed, metabolized, and excreted by the body, as well as what the medicines do to the body. There is also the ethical requirement to minimize both the number of studies in children and the number of children recruited to studies (see [6] and [7]). In many cases, children can be more vulnerable to the effects of a drug than people in older age groups. Children are also not able to consent for themselves, while many parents are reluctant to risk having their children randomized to treatment assignment. On the logistical side, there might be no active control available for pediatric use while withholding treatment may pose risk of serious harm, or the disease may have low incidence in children. Regardless of the difficulty of patient enrollment, drugs for children are typically subject to the same statistical evidentiary standards for efficacy and safety as drugs in adults, which translates to enrolling a sufficient number of patients. The number of patients that can be enrolled may necessitate a

more innovative statistical design, or require multi-site or even global studies to accrue sufficient patients.

One way to meet sample size and efficacy requirements is to *extrapolate* information from data about the drug and the condition in older age groups. Extrapolation is to extend information and conclusions available from studies in one or more subgroups of the patient population (the *source* population), to make inferences for another subgroup of the population (the *target* population), thus reducing the need to generate additional information to reach conclusions for the target population, or condition or medicinal product. Extrapolation may be needed to avoid additional studies in the target population for ethical reasons, feasibility restrictions, or efficiency, to allow resources be allocated to areas where studies are the most needed. The European Medicines Agency (EMA) issued a concept paper on extrapolation of efficacy and safety in medicinal development in May 2013 [8]. The paper proposed a framework for an explicit and systematic approach which sets out when, to what extent, and how extrapolation can be applied in drug development. Recently, the EMA issued a reflection paper that uniquely addresses extrapolation in the context of pediatric drug development [9]. In particular, it recommends a systematic review to describe the mechanisms and characterize differences between the source and target population on medicine disposition and effects, disease manifestation and progression, and clinical response to treatment.

Extrapolation has been previously utilized in the pediatric trial setting. Mulugeta et al. [10] provided a systematic review of approaches used for matching adult systemic exposures as the basis for dose selection in pediatric trials submitted to the US FDA between 1998 to 2012. Dunne et al. [11] described two categories of extrapolation: *full* (also referred to as *complete*) extrapolation, where adult data are used directly to establish pediatric device safety or efficacy, and *partial* extrapolation, where adult data are statistically combined with pediatric data to make such determinations. Full extrapolation relies on robust data supporting the assumptions that there are similar disease progressions, responses to intervention, and exposure-response relationships in adult and pediatric populations. In this setting, the effective dose can be identified by matching systemic exposures between adult and pediatric populations. Consequently pediatric development supported by pediatric pharmacokinetic and safety data or, in certain cases, pediatric safety data only, can be considered as adequate. Partial extrapolation of efficacy is used when there is uncertainty about at least one of the assumptions underlying complete extrapolation as stated above. In this category, the pediatric development strategy ranges from a single adequate, well-controlled trial to confirm efficacy, to a pharmacokinetic/pharmacodynamic (PK/PD, also known as exposure-response) study to confirm response in the pediatric population.

Based on these definitions, Dunne et al. [11] reviewed 370 pediatric studies (166 products) submitted to FDA from 1998 to 2008, and found that 82.5% of the drug products utilized extrapolation, including 14.5% using complete extrapolation and 68% using partial extrapolation. Sachs [12] summarized some real experiences of pediatric extrapolation during a pediatric investigator training session; studies of Denavir cream (penciclovir) for treatment of recurrent herpes labialis (cold sores) and Zmax oral suspension (azithromycin) for treatment of community-acquired pneumonia used full extrapolation of efficacy. Studies of Maxalt and Maxalt-MLT (rizatriptan benzoate) for treatment of migraine, Viread (tenofovir disoproxil fumarate) for treatment of chronic Hepatitis B, Aloxi (palonesetron) for prevention of acute nausea and vomiting associated with chemotherapy, Nitropress (sodium nitroprusside) for immediate reduction of blood pressure in hypertensive crisis, and Sabril (vigabitrin) as adjunctive therapy

for refractory complex partial seizures have utilized partial extrapolation in pediatrics.

As defined, central to extrapolation is the exposure/response relationship, which needs to be developed in adults if borrowing to pediatric population is to be done. However, current methods are often ad hoc and depend crucially on knowing the appropriate amount of information to borrow from the adult data. In fact, the EMA concept paper [8] pointed out some key issues in developing an extrapolation framework including how to define and quantify similarity of disease progression, PK/PD, or clinical response to treatment between adult and pediatric populations; how to weigh the strength of prior information; how to quantify the uncertainty of extrapolation assumptions; and how to validate assumptions. All these issues call for novel quantitative extrapolation approaches to enrich our methodological toolkit.

Recent years have seen an increase in the use of Bayesian statistical methods (see e.g. [13]) in the design, interim monitoring, and final analysis of clinical trials, especially for early phase studies. By offering a formal statistical framework for incorporating all sources of knowledge (structural constraints, expert opinion, and both historical and experimental data), these methods offer the possibility of a substantially reduced sample size, thanks to their more efficient use of information [14]. This in turn typically leads to increase in statistical power, reductions in cost and ethical hazard, the latter since fewer patients need be exposed to inferior treatments.

More than a decade ago, Goodman and Sladky [15] used a case study to demonstrate how a Bayesian method could be used to incorporate prior information on treatment efficacy from adults to design a randomized non-inferiority trial of IVIg (intravenous immune globulin) vs. plasmapheresis in children. They utilized a Bayesian normal errors model for the hazard ratio of time to independent walking. The resulting sample size required for the pediatric study was smaller than using traditional "frequentist" statistical methods, due to borrowing strength from adult information. (We hasten to add that, while the methods use of adult information was restricted to their scale – and not to their mean, the use of which was feared might "bias the results" – the design was at that time viewed as too controversial, and the trial was never run.) The Bayesian approach also allows quantitative definition of the similarity of disease progression, PK/PD, or clinical response to treatment between adult and pediatric populations. Schoenfeld et al. [16] proposed a Bayesian approach to allow borrowing strength from previous or simultaneous adult trials in designing a pediatric efficacy trial, which is a quantitative partial extrapolation. The authors applied a hierarchical model in which the efficacy parameters from the adult trial and the pediatric trial are considered to be draws from a common normal distribution, where its mean has a non-informative prior. The variance of this distribution measures the variation between the response of adults and children to a treatment, and subsequently connects this variance with an *effective sample size*, representing a discounted amount of information borrowed from the adult study. Furthermore, by specifying a prior, one can also quantify the uncertainty of extrapolation assumptions. Gastonguay [17] described various Bayesian methods in PK pediatric studies through modeling and simulations. The variability of the PK parameters from adult studies was used for pediatric studies, with age factored in as a random effect. This approach makes use of adult data and different age groups in pediatric data for guiding simulations in the design stage. The author also proposed a Bayesian decision tree type of dosing, where adult data was used and scaled down by age and weight, and then combined with data from the current trial to predict dosing. The authors argued that this method can improve the precision of PK parameter estimation. Recently, Wadsworth et al. [18] provided a systematic review of statistical methods that could be used to optimize extrapolation

Study Phase	Drug/ Device	Main/ Supportive Analysis	Blinded	Endpoint	Prior Used	Sample Size	PIP/PSP Agreed	Reference Reference
III	Drug	Main	No	Disease remission within 6 months of randomization	MYCYC trial Expert opinions	40	No, investigator Sponsored	MYPAN trial
III	Drug	Main		Hazard ratio of time to independent walking				GBS trial [15]
I (PK)	Drug			PK model with age as as a variable				AMD trial [19]

Table 1: Examples of use of Bayesian statistics in pediatric trials.

in pediatric drug development programmes. The review identified 58 Bayesian methods from 25 papers. Of these, 54 methods sought to create an informative prior.

1.2 Barriers to the Use of Bayesian Statistics in Pediatric Trials

Examples of current use of Bayesian statistics in pediatric clinical trials are limited. We searched for available examples from industry and academic environment and available publications. The results are presented in Table 1.

To date, Bayesian methods have seen only limited use in industry-sponsored pediatric trials, with applicants often reluctant to use them due to lack of in-house expertise, high computational requirements, and nervousness regarding the anticipated reaction of regulatory authorities. In fact, Bayesian borrowing of strength from auxiliary data has long been encouraged in the case of medical device trials by the Center for Devices and Radiological Health (CDRH) at the U.S. Food and Drug Administration (FDA) (c.f. [20]), where sample sizes are small but reliable historical data are often plentiful [14]. A recently published FDA guidance document ([21]; c.f. [22]), co-signed by CDRH and the Center for Biologics Evaluation and Research (CBER), helps to clarify the situations under which full or partial extrapolation are justified in device trials, and the Bayesian and frequentist statistical methods that might be used. Adoption of Bayesian methods in the development of medicinal products by sponsors and regulators has been slower, largely due to concerns that formal use of information external to the trial might inflate the trial's Type I error rate. Some of the concerns stem from the generally perceived view that applicants of the Bayesian methods will "cherry-pick" datasets favorable to their position, discarding unfavorable data. None-the-less, there is an increasing willingness for the use of Bayesian methods for data augmentation, as the need for more efficient methods to handle pediatric, orphan, and other hard-to-study diseases increases.

Among biostatisticians working in later phase drug trials, the working group observes that reluctance to use Bayesian methods appears to have three primary causes. First, the Bayesian approach does require an initial assessment of the commensurability of the various sources of information, which is often difficult for investigators to make. Data may be available from earlier trials of a similar drug, or on related but not identical patient populations, but the precise amount of trust one should place in this information may be unclear. If the auxiliary and primary data sources conflict, the result can be a higher than expected Type I error (false positive) rate, as well as the possibility of a costlier and lengthier trial, since extra experimental information will be needed to resolve the conflict. The second cause is the paucity of appropriate and sufficiently user-friendly, robust, and well-documented software for Bayesian clinical trials. While Bayesian software programs are beginning to emerge, such software typically relies on

Markov chain Monte Carlo (MCMC) or other complex computational algorithms that can be difficult for inexperienced users to implement and monitor. Recent computational advances, however, permit routine summarization of key model outputs (say, the probability a drug is effective and/or safe given all evidence collected to date). These successes suggest the potential for broader application to pediatric disease settings. A third cause for reluctance to use Bayesian approaches in medical product development is operational issues, including (but not limited to) a limited number of trained Bayesian statisticians, or the time required to perform and check the proposed Bayesian simulation study and data analysis.

1.3 Opportunities for Bayesian Approaches in Pediatric Drug Development

Because new approaches are needed for ethical reasons and to address feasibility, there may be opportunities for greater use of Bayesian and other alternative statistical methods in pediatric research. Indeed, children represent a large under-served population of "therapeutic orphans", since an estimated 80% of children are treated "off-label", i.e., physicians use their professional judgment, based on extrapolation of information from adult data, to prescribe a drug to children. This means the safety, efficacy, and PK/PD of such drug therapies in children is unknown. It is understandable, however, that the decision on the use of Bayesian methods should be made on an individual, case-by-case basis, since development programs differ for each medicinal product.

Apart from medicinal products developed specifically for children, industry is encouraged by regional regulation to investigate their medicines in children, even though certain waivers apply and some medicinal products are excluded from this requirement. In those cases, investigation in children starts only after initial proof of medicinal product safety and efficacy has been gathered in adults. Based on the information collected and disease in question, appropriate models are used to properly design a pediatric program and to predict the starting dose in children. Increased efforts could be made to more efficiently use relevant historical data, both adult and pediatric.

In our view, an optimal pediatric development plan that employs the Bayesian approach must consider the following features:

- Understanding the disease in question, the similarity between adults and children and its incidence and prevalence in children;
- Understanding the expected treatment response differences in children, if any;
- Allowing the possibility of borrowing information from previous studies (adult and pediatric) and specifying the proper extent of this borrowing, perhaps as determined by study quality or the similarity of the various data sources and expert opinion;
- Optimizing trial sample size through better use of prior data while maintaining adequate study power and monitoring Type I error rate.

The objective of this paper is to develop, illustrate, and provide suggestions on statistical methods that could be used by pharmaceutical industry and CRO biostatisticians to efficiently design pediatric studies. Our approaches should also benefit investigators considering new medicinal products for children, by offering innovative methods that can improve understanding of treatment effects and advance the pace of evidence collected. Use of these approaches should

lead to improved pediatric development programs that use available information in the most efficient manner, and allow for faster and more robust conclusions regarding a product's pediatric efficacy and safety. In addition, the suggestions presented here take into consideration the agency-published guidance on clinical trials in small populations [23], the Pediatric Medical Device Safety & Improvement Act of 2007 and PREA [3], and the guidance on "Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices" [21].

2 Bayesian Statistics

2.1 Background

Conventional frequentist statistical approaches to evaluating medicinal products' efficacy and safety require relatively large numbers of subjects, with each investigational medicinal product requiring its own sequence of costly clinical trials. As mentioned above, Bayesian statistical methods permit the combining of information from disparate sources.

To establish a bit of notation, let D be the observed data arising from a sampling distribution $F(\cdot|\theta)$. Suppose $\theta = g(\vartheta)$ is the target parameter and g is the link function needed to transform θ to the mean response scale (e.g., the logistic function in the case of a binary response). Then the likelihood of θ given D is denoted by $L(\theta|D)$. Rather than simply maximize this function to find a point estimate of θ, Bayesian methods instead add a *prior* distribution $\pi(\theta)$ summarizing what is known about the parameter θ before the data D are collected. The prior may be based on auxiliary information (such as expert clinical opinion), or taken as "non-informative", so that the data D will dominate the inference. A Bayesian then computes the *posterior* distribution $p(\theta|D)$ via Bayes Rule [24] using either probability calculus or, where closed form calculations are not possible, a numerical approach such as Markov chain Monte Carlo (MCMC; see e.g. [13], Ch. 4). In the case when the prior is the identity function ($\pi(\theta) \propto 1$; a so-called "flat prior"), the prior vanishes from this calculation, and the posterior distribution coincides with the renormalized frequentist likelihood function. In any case, all statistical inference follows straightforwardly from the posterior distribution. For example, the "middle" (mean or median) of the posterior is a sensible point estimate for θ, and the 2.5 and 97.5 percentiles of this distribution offer a sensible 95% confidence interval (known as a *credible set* in Bayesian parlance).

Hypothesis testing for Bayesians historically proceeded using not p-values, which are anathema to Bayesians due to their failure to obey the Likelihood Principle [25], but by Bayes factors (the ratio of posterior to prior odds in favor of one of the two hypotheses; essentially a measure of the adjusted odds given by the data alone). However, since Bayes factors can be difficult to compute and are not well-defined when the prior distribution is fully non-informative, in recent years Bayesians have turned to more general approaches, including the use of decision criteria based on the posterior probability, such as the Deviance Information Criterion (DIC; see [26]), a penalized likelihood criterion and generalization of the Akaike Information Criterion (AIC) that is the sum of two terms, one capturing model fit and the other a penalty for model complexity.

The simultaneous emergence of MCMC methods such as the Gibbs sampler and high-speed desktop computing in the 1990s has led to an explosion in applied Bayesian work, much of it implemented using the BUGS software package [27]. In addition, a large amount of Bayesian analyses are also now performed in R, SAS, and other mainstream statistical packages, as well as specialty software freely available from the CRAN archive (`http://cran.r-project.org/`).

2.2 General Methods for Extrapolation

2.2.1 Two-step Approach

Let $\boldsymbol{D}_0 = \{Y_1, \ldots, Y_K\}$ be the data from a set of historical trials for a drug approved for an indication in the adult population. These data are then used as supplemental data to determine whether the drug's indication can be extended to the pediatric population whose available primary data is denoted by $D = \{Y_{K+1}\}$. Suppose ϑ_{K+1} is the associated parameter for the primary data, which is analogous to the set of parameters $\boldsymbol{\vartheta} = \{\vartheta_1, \ldots, \vartheta_K\}$ from the supplemental data. If F is the sampling distribution in the supplemental and primary data, and the instances with which the ϑ_k, $k = 1, \ldots, K+1$, occur are indistinguishable (i.e., the ϑ_k are *exchangeable*), then there exists a distribution G such that the transformed parameters $\theta_k = g(\vartheta_k)$ are a random sample from G. This can be hierarchically represented as

$$Y_k|\vartheta_k \sim F(\vartheta_k; n_k),\ \theta_k = g(\vartheta_k)|\zeta \sim G(\zeta),\ \zeta \sim H\ , \tag{1}$$

where the second stage prior distribution (or *hyperprior*) H describes the initially available information about ζ [28], and n_k is the sample size associated with Y_k. For simplicity, suppose the prior (random effects) distribution G is $\mathrm{N}(\theta, \tau^2)$ with $\zeta = (\theta, \tau^2)$. That is, the exchangeability of the parameters in the supplemental data does not mean that $\theta_1 = \cdots = \theta_K$, but rather suggests homogeneity, in that they have a common mean θ and between-trial standard deviation τ that can be conceptualized as

$$\theta_1, \ldots, \theta_K|\theta, \tau^2 \sim \mathrm{N}(\theta, \tau^2)\ , \tag{2}$$

where the hyperprior distribution of the hyperparameters H is $\pi(\zeta) = \pi(\theta|\tau^2)\pi(\tau^2)$. Note that a more flexible random effects distribution for $\theta_1, \ldots, \theta_K$, such as a t-distribution or Dirichlet process mixture (c.f. [29]), can also be considered. Assessment of exchangeability of the ϑ's is fundamental. An important consequence of exchangeability is "borrowing of strength," or combination of information. The information about the similarity of the θ_k informs inferences about the distribution of the θ_k across groups. Thus, inferences for the group-level parameters θ_k reflect not just the information in Y_k itself, but, via the hierarchical model, will also draw on relevant information in the other groups. If there is dissimilarity in relevant factors (e.g., patient population, patient management or standards of care, operator training/experience, etc.) between studies that could lead us to believe that the assumption of a common random effect distribution may not be valid, then borrowing of information is not warranted and the model adjusts by inflating our estimate of τ^2.

From the specification above, the likelihood function given the supplemental data is $L(\boldsymbol{\theta}, \theta, \tau^2|\boldsymbol{D}_0) = \prod_{k=1}^{K} f(Y_k|n_k, \theta_k, \theta, \tau^2)$, where f is the density of F and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$. Then the joint posterior distribution is $p(\boldsymbol{\theta}, \theta, \tau^2|\boldsymbol{D}_0) \propto L(\boldsymbol{\theta}, \theta, \tau^2|\boldsymbol{D}_0)\pi(\boldsymbol{\theta}|\theta)\pi(\theta|\tau^2)\pi(\tau^2)$. Integrating out $\boldsymbol{\theta}$ and τ^2, a prior for θ_{K+1} can be specified as

$$\pi(\theta_{K+1}) \propto p(\theta|\boldsymbol{D}_0) = \int\int L(\boldsymbol{\theta}, \theta, \tau^2|\boldsymbol{D}_0)\pi(\boldsymbol{\theta}|\theta)\pi(\theta|\tau^2)\pi(\tau^2)d\boldsymbol{\theta}d\tau^2\ . \tag{3}$$

In simple cases, this marginal posterior has a closed form, so that a straightforward application of Bayes Rule yields the desired posterior $p(\theta_{K+1}|\boldsymbol{D}_0) \propto L(D|\theta_{K+1})\pi(\theta_{K+1})$. For example, when the likelihood and priors on $\boldsymbol{\theta}$ and θ_{K+1} are normal and the prior for τ^2 is inverse gamma, the marginal posterior in (3) emerges as a Student's t distribution. However, in many cases, (3) does

not lead to a kernel that is recognizably associated with a standard distribution. This motivates and MCMC-based approximation $\hat{p}(\theta|\tau^2, \boldsymbol{\theta}, \boldsymbol{D}_0)$. Schmidli et al. [30] suggest approximating the marginal posterior as a finite mixture of prior densities [31],

$$\hat{p}(\theta|\boldsymbol{D}_0) = \sum_{i=1}^{m} p_i \phi(\theta|\theta_i, \tau_i^2), \quad p_i > 0, \quad \sum_{i=1}^{m} p_i = 1 \tag{4}$$

with the value of m chosen to be small, where ϕ is a density of the same kernel as the density of G. The selection of the number of components is guided by a measure of divergence so that the approximation is close to $p(\theta|D)$ [32]. In addition, robustness of the prior can be enhanced by adding a vague distribution $p(\theta)$, i.e.,

$$\hat{\pi}(\theta_{K+1}) = (1-w)\hat{p}(\theta|\boldsymbol{D}_0) + wp(\theta), \tag{5}$$

where $0 < w < 1$. Alternatively, one can approximate this posterior by a mixture of normals where the mixing distribution is a Dirichlet Process (see Ferguson [33]). In general, heavy-tailed distributions have been shown to produce more reasonable conflict resolution, typically by favoring one source of information over another. The less favored source is wholly or partially rejected as the conflict becomes increasingly extreme [34]. Finally, given this prior, the desired posterior arises as $p(\theta_{K+1}|D, \boldsymbol{D}_0) \propto L(\theta_{K+1}|D)\hat{\pi}(\theta_{K+1})$.

This approach can be readily extended to incorporate individual-level data. Here the random effects model needs to be extended to incorporate another layer in the hierarchy such that the parameters for each individual are exchangeable within studies, but not across studies. The overall means from each study, though, are still assumed exchangeable to facilitate borrowing between studies. Covariates can be added into the model as well, which changes the parameter θ to be vector valued and τ^2 into a covariance matrix. In particular, the extension to include two-arm studies is straightforward, as the likelihood of the primary data can be written as $L(\theta + t_{ik}\lambda|D)$, where t_{ik} is a treatment indicator for subject i in study k. Here, interest typically lies in the overall treatment difference λ, which is ordinarily assigned a non-informative prior. Lastly, the two-step approach is practical because the prior is relevant not only in the analysis of the new data, but in planning or designing the new trial itself, which includes the evaluation of frequentist operating characteristics (power, Type I error, etc.) necessary for a good statistical plan.

2.2.2 Combined Approaches

While the two-step approach calculates the posterior of the adult supplemental data before combining it with the likelihood of the primary data, a combined data approach instead combines all the data in a single hierarchical model. Technically, the two approaches are equivalent due to exchangeability, but if "robustification" as in (5) is involved then they are not. Here we may begin by assuming that $\theta_1 = \ldots = \theta_K = \theta_{K+1} = \theta$, instead of putting a homogeneity constraint, τ^2, on the θ_ks. We then obtain the posterior $p(\theta|D, \boldsymbol{D}_0)$ as proportional to $L(\theta|D)L(\theta|\boldsymbol{D}_0)\pi(\theta)$. In addition, unlike the two-step approach where to accommodate prior-data conflict the posterior $p(\theta|\tau^2, \boldsymbol{\theta}, \boldsymbol{D}_0)$ is adjusted to be heavy-tailed, the combined approach modifies the joint likelihood instead. One such modification is to downweight the likelihood of the adult data through a *power prior* or a *commensurate prior*. The power prior approach [35] assumes that $\theta_1 = \ldots = \theta_{K+1} = \theta$,

but downweights the supplemental (adult) likelihood by raising it to a power α_0 that is between 0 and 1. The posterior distribution then arises as

$$p(\theta|D, \boldsymbol{D}_0, \alpha_0) \propto L(\theta|D)L(\theta|\boldsymbol{D}_0)^{a_0}\pi(\theta). \tag{6}$$

Note that a_0 controls how much information will be borrowed from the auxiliary (adult) data to supplement the (fully-utilized) primary data; e.g., $a_0 = 1$ means full borrowing, while $a_0 = 0$ implies no borrowing. Such control is important in cases where there is heterogeneity between the supplemental and primary data, or when equal weighting of primary and supplemental samples is inappropriate. In fact, if the power priors are fixed, there is a one-to-one relationship between the power parameter of the power prior and the variance (equivalently, the effective sample size; see Section 2.3 below) of the prior; the relationship is particularly straightforward in the normal likelihood setting.

While a fixed power prior can be useful when data from historical trials are believed to be equally reliable, in other settings it is more reasonable to use study-specific weights. The estimates of these weights can be data-driven by assuming that study-specific a_{0k} are independently distributed as Beta(b_1, b_2) random variables, where $b_1 > 0$ and $b_2 > 0$ are predetermined constants. This gives the joint power prior

$$\pi(\theta, \boldsymbol{a}_0|\boldsymbol{D}_0) \propto \left[\prod_{k=1}^{K} f(Y_k|\theta)^{a_{0,k}}\right] \pi(\theta)\pi(\boldsymbol{a}_0)\ , \tag{7}$$

where $\boldsymbol{a}_0 = (a_{0,1}, \ldots, a_{0,K})'$, $\pi(a_{0,k}) = \text{Beta}(b_1, b_2)$, and the joint posterior is obtained as $p(\theta, \boldsymbol{a}_0|D, \boldsymbol{D}_0) \propto L(\theta|D)\pi(\theta, \boldsymbol{a}_0|\boldsymbol{D}_0)$. Other modifications to the prior specification of the power $\boldsymbol{a}_0$, include use of time-dependent $a_{0,k}$, or ordering the studies by insisting $a_{0,1} < a_{0,2} < \cdots < a_{0,K}$ and putting an ordered Dirichlet distribution as a prior [36].

Note that in (7), the joint power prior $\pi(\theta, \boldsymbol{a}_0|\boldsymbol{D}_0)$ is improper and inappropriately scaled, since it fails to account for the normalizing constant $A(\boldsymbol{a}_0) \equiv \int \prod_{k=1}^{K} f(Y_k|\theta)^{a_{0,k}}\pi(\theta)d\theta$, which is a function of the unknown power parameter vector $\boldsymbol{a}_0$. Several authors have spotted this problem; Duan et al. [37] refer to the version that *does* properly include this normalizing constant, $\pi(\theta, \boldsymbol{a}_0|\boldsymbol{D}_0) \propto A(\boldsymbol{a}_0)^{-1}\left[\prod_{k=1}^{K} f(Y_k|\theta)^{a_{0,k}}\right]\pi(\theta)\pi(\boldsymbol{a}_0)$, as the *modified power prior*. This prior *is* proper and appropriately scaled over $\mathcal{A} = \{\boldsymbol{a}_0 : 0 < A(\boldsymbol{a}_0) < \infty\}$. In cases where $f(Y_k|\theta)$ and $\pi(\theta)$ are a conjugate pair, the calculation of the modified power prior is straightforward. For instance, when $f(Y_k; n_k, \theta) = \text{Binomial}(n_k, \theta)$ with $\theta \sim \text{Beta}(\kappa\mu, \kappa(1-\mu))$, we have $\prod_{k=1}^{K} f(Y_k|\theta)^{a_{0,k}} \propto \prod_{k=1}^{K} \theta^{n_k a_{0,k}}(1-\theta)^{(n_k - Y_k)a_{0,k}}$, which is proportional to another beta distribution. Hence, the modified power prior $\pi(\theta, \boldsymbol{a}_0|\boldsymbol{D}_0)$ arises as a Beta$(\kappa^*\mu^*, \kappa^*(1-\mu^*))$, where $\kappa^* = \kappa + \sum_{k=1}^{K} n_k a_{0,k}$ is the updated sample size and $\mu^* = \kappa\mu/\kappa^* + \hat{\theta}\sum_{k=1}^{K} n_k a_{0,k}/\kappa^*$ is the posterior mean of θ expressed as the weighted average of μ, the prior mean of θ, and the weighted sample mean $\hat{\theta} = \sum_{k=1}^{K} Y_k a_{0,k} / \sum_{k=1}^{K} n_k a_{0,k}$. Given this modified power prior, the joint posterior of $p(\theta, \boldsymbol{a_0}|D, \boldsymbol{D}_0)$ is proportional to Beta$(\tilde{\kappa}\tilde{\mu}, \tilde{\kappa}(1-\tilde{\mu}))\pi(\boldsymbol{a}_0)$ where $\tilde{\kappa} = \kappa + n_{K+1} + \sum_{k=1}^{K} n_k a_{0,k}$ and $\tilde{\mu} = \tilde{\kappa}^{-1}(\kappa\mu + Y_{K+1} + \sum_{k=1}^{K} Y_k a_{0,k})$. Note that, similar to above, $\tilde{\mu}$ can be written as a weighted average of the prior mean μ and the weighted combined sample mean $\tilde{\theta} = [Y_{K+1} + \sum_{k=1}^{K} Y_k a_{0,k}]/[n_{K+1} + \sum_{k=1}^{K} n_k a_{0,k}]$, where of course the pediatric data Y_{K+1} get full weight. Under a flat prior $\pi(\boldsymbol{a}_0) = 1$, we obtain $p(\theta, \boldsymbol{a_0}|D, \boldsymbol{D}_0) \propto \text{Beta}(\tilde{\kappa}\tilde{\mu}, \tilde{\kappa}(1-\tilde{\mu}))$, and one can use MCMC techniques to determine the posterior mean of this density over a unit cube in $(K+1)$ dimensions, which corresponds to the posterior mean of θ and $a_{0,k}$, $k = 1, \ldots, K$.

Because of the difficulty in using power priors that adapt to the heterogeneity between supplementary and primary data, there are recent proposals to estimate these powers using propensity scores (e.g. Zhao et al. [38]). When individual patient data are available, the propensity score $e_{ik} = e(\mathbf{X}_{ik})$ for subject i within study k with vector $\mathbf{X}_{ik}$ of observed covariates is the conditional probability of having been included in the pediatric data ($Z = 1$), including those subjects that are already in the pediatric data, instead of one of the other datasets ($Z = 0$) given observed pre-treatment characteristics $\mathbf{X}_{ik}$; then $e(\mathbf{X}_{ik}) = \Pr(Z_{ik} = 1|\mathbf{X}_{ik})$. One can then use the e_{ik} as the weights $a_{0,ik}$. Other proposals suggest using the correlation of exposures in adults and pediatrics. These modifications have yet to be investigated thoroughly.

By contrast, the *commensurate prior* approach [39, 40] modifies the likelihood by the extent to which a parameter in the pediatric trial varies about the analogous parameter in a set of supplementary adult trials when the direction of the bias is unknown. In the case of (2), the approach sets $\theta_1 = \cdots = \theta_K = \theta_0$; i.e., the adult data parameters are constrained to all be equal to θ_0, with $\theta_{K+1} \neq \theta_0$ and $Var(\theta_{K+1}|\theta_0) = \eta^{-1}$. The commensurate prior approach then specifies a hierarchical model with posterior

$$p(\theta_{K+1}, \theta_0, \eta|D, \boldsymbol{D}_0) \propto L(\theta_{K+1}|D)L(\theta_0|\boldsymbol{D}_0)\pi(\theta_{K+1}|\theta_0, \eta)\pi(\theta_0)\pi(\eta) \; . \tag{8}$$

The $\pi(\theta_{K+1}|\theta_0, \eta)$ component of this expression is referred to as the commensurate prior, where the *commensurability parameter* η is inversely related to the between-study variance parameter for the random-effects meta-analytic models, and $\pi(\theta_0)$ reflects (typically vague) initial prior knowledge about θ_0 before seeing $\boldsymbol{D}_0$. While any distribution over the positive reals could be used, this crucial parameter η is often assigned a "spike and slab" hyperprior, which is a mixture of a point mass at some large value R and a uniform "slab" of positive values near 0. For instance, referring to the example described previously where the observations have binomial sampling distributions, the logistic link function first transforms the expectations of $Y_1, \ldots, Y_K, Y_{K+1}$ such that $\log(p_k/(1-p_k)) = \theta_0$, $k = 1, \ldots, K$, and $\log(p_{K+1}/(1-p_{K+1})) = \theta_{K+1}$. Thus the commensurate prior (8) is proportional to $\prod_{k=1}^{K} g^{-1}(\theta_0)^{Y_k}(1-g^{-1}(\theta_0))^{n_k-Y_k}\mathrm{N}(\theta_{K+1}|\theta_0, \eta)\pi(\theta_0)\pi(\eta)$. A normal commensurate prior is usually used as it most conveniently captures between-study variability through η and its hyperprior, although the Cauchy prior has also been recommended [41]. Note here that as $\eta \to \infty$, $\theta_{K+1} \to \theta_0$ and the prior assumes full commensurability of the pediatric and adult data; for η close to 0, commensurability is rejected and the impact of the adult data is diminished. In standard parametric cases, implementation of commensurate prior analysis is straightforward, e.g. in the `BUGS` language [27]; in more advanced settings, specialized algorithms may be required [42, 43].

Still, using any of the approaches in this section, the analytical computations can be laborious. Simplifications arising from using conjugate pairs can help reduce the computational burden. Limited noncommercial software is beginning to become available for implementing these approaches, particularly for 2-arm studies featuring existence of supplemental data on the control treatment, concurrent data for the same control treatment, and current data on the experimental treatment; see e.g. `http://research.mdacc.tmc.edu/SmeeactWeb/`.

2.2.3 Partial Exchangeability Approach

As frequently mentioned previously, there is trepidation that the adult supplemental data and the pediatric data maybe dissimilar, making data combination inappropriate. Another suggestion

to mitigate this concern is to add a new layer to the hierarchical model (1) that expresses non-exchangeability (see [44, 22]), i.e.,

$$Y_{jk}|\vartheta_{jk} \sim F(\vartheta_{jk}; n_{jk}),\ \theta_{jk} = g(\vartheta_{jk})|\zeta_j \sim G(\zeta_j),\ \zeta_j|\gamma \sim H(\gamma),\ \gamma \sim P. \tag{9}$$

Here, Y_{jk} denotes the observation from the kth study within the jth population (adult or pediatric) with parameter ϑ_{jk}. Conditional on the population covariate level or within each population j, the studies are exchangeable, but not across populations. To facilitate borrowing between the adult and pediatric studies, they are connected by assuming their exchangeability. The net effect is a *partially exchangeable* model. Thus we are interested in the posterior

$$p(\zeta_{\text{peds}}, \boldsymbol{\theta}, \zeta_{\text{adult}}, \gamma, |D, \boldsymbol{D}_0) \propto L(\theta_{\text{peds}}|D)L(\boldsymbol{\theta}_{\text{adult}}|\boldsymbol{D}_0)\pi(\boldsymbol{\theta})\pi(\zeta_{\text{peds}})\pi(\zeta_{\text{adult}})\pi(\gamma), \tag{10}$$

where $\boldsymbol{\theta} = (\theta_{\text{peds}}, \boldsymbol{\theta}_{\text{adult}})$. This can be decomposed into conditional posteriors for convenient implementation via MCMC.

2.3 Effective Sample Size

While our attention has been focused on the incorporation of well-chosen priors based on supplemental data, we have yet to formally quantify the amount of information borrowed from the adult data through the prior. One intuitive tool for understanding how much information is gained is the *effective sample size* (ESS). With conjugate priors for the exponential family, the ESS is easily obtained [32] as illustrated in the example mentioned in Subsection 2.2.2. For example, if the sampling distribution for the primary data $f(Y_{K+1}; n_{K+1})$ is Binomial(n_{K+1}, θ) and the beta prior based on supplemental data $\pi(\theta, \boldsymbol{a}_0|\boldsymbol{D}_0)$ is Beta$(\kappa^*\mu^*, \kappa^*(1-\mu^*))$ where $\kappa^* = \kappa + \sum_{k=1}^K n_k a_{0,k}$ and $\mu^* = \kappa\mu/\kappa^* + \hat{\theta}\sum_{k=1}^K n_k a_{0,k}/\kappa^*$, then the posterior is proportional to Beta$(\tilde{\kappa}\tilde{\mu}, \tilde{\kappa}(1-\tilde{\mu}))\pi(\boldsymbol{a}_0)$ where $\tilde{\kappa} = \kappa + n_{K+1} + \sum_{k=1}^K n_k a_{0,k}$ and $\tilde{\mu} = \tilde{\kappa}^{-1}(\kappa\mu + Y_{K+1} + \sum_{k=1}^K Y_k a_{0,k})$. Hence, the total effective sample size is $\kappa + n_{K+1} + \sum_{k=1}^K n_k a_{0,k}$, which reflects a gain of $\kappa + \sum_{k=1}^K n_k a_{0,k}$ effective patients from both the initial prior and the adult data. Based on this observation, Morita et al. [45] proposed a definition for the ESS as the interpolated value of k that minimizes this "prior-to-posterior" distance. In particular, by constructing an "ε-information" prior $p_0(\theta)$, considering a hypothetical sample $\boldsymbol{D}_0$ of size k and the posterior $p_k(\theta|\boldsymbol{D}_0)$, and computing a distance between $p_k(\theta|\boldsymbol{D}_0)$ and $p_0(\theta)$ in terms of the curvature (second derivatives) of $\log(p_0(\theta))$ and $\log(p_k(\theta|\boldsymbol{D}_0))$, the value of k minimizing the distance is the prior ESS. An alternative heuristic formula introduced by Malec [46] is to calculate ESS based on the ratio of the posterior variance in the study without borrowing from prior studies, to the posterior variance in the study with borrowing, multiplied by the study sample size. Indeed, in the case of Gaussian likelihoods and priors, such variance-based definitions of ESS yield intuitive results in closed form, similar to the beta-binomial case.

2.4 MCMC Convergence Diagnostics and Model Fit

When estimates of posterior means are obtained via MCMC, assessing convergence through diagnostics (e.g. Gelman and Rubin [47] and Raftery and Lewis [48]; c.f.Carlin and Louis [13] for a review) is imperative. Many of these diagnostics are implemented directly within `BUGS`, or within the CODA [49] library in `R`. In addition, plotting multiple chains also helps visualize mixing of

chains to a stationary distribution. Significant positive serial autocorrelation may suggest the need for longer chains or a more efficient algorithm. The effective number of simulation draws, which is provided as n_{eff} in `Stan` output [50] and similarly obtained in `BUGS` or `JAGS` [51] should equal the actual number of posterior draws if there is essentially no autocorrelation in the chains. Lastly, sensitivity of the results to the assumed sampling distribution, functional form of parameter relationships, prior/hyperprior specification, initial values, and other model and algorithm characteristics should also be checked.

For assessing model adequacy, two relative measures to compare models are often calculated: the Deviance Information Criterion (DIC) and the log-pseudo marginal likelihood (LPML). The DIC is a Bayesian hierarchical modeling extension of the Akaike Information Criterion (AIC) that combines an assessment of classical "plug-in" measure of fit and a Bayesian measure of model complexity. It is implemented in `BUGS` and `JAGS`, and calculated as $D(\bar{\boldsymbol{\vartheta}}) + 2p_D$, where $D(\bar{\boldsymbol{\vartheta}})$ is the deviance evaluated at the posterior mean of the parameters, $\bar{\boldsymbol{\vartheta}}$, and p_D is "effective number of parameters," which operates as a complexity penalty (see [26]). A small value of the DIC is desired, with changes of less than 5-10 units often thought of as not worthy of mention (i.e., the two models are equally adequate).

Another way to evaluate a model is through the accuracy of its predictions, as captured by the LPML. This is a leave-one-out cross validation measure based on predictive densities [52]. While this estimate is efficient, it can be unstable; see Gelfand and Dey [53] for alternatives. Since the LPML is a leave-one-out cross-validation method, it can be interpreted as a predictive measure for a future replicate of a given data set. The larger its value, the better the fit of the data to the prescribed model. The LPML is also useful when comparing models having different likelihoods (say, normal versus Cauchy) or models with discrete parameters, since the plug-in deviance required by the p_D term in DIC may not be available when the posterior mean of a discrete parameter is either undefined or not guaranteed to be inside the discrete parameter space [27].

2.5 Operating Characteristics

Like therapeutic drugs developed for adult indications, drugs seeking pediatric labeling are required to provide sufficient evidence to support safe and effective use [2, 3, 54]. One interpretation of this requirement is that regulatory agencies can request that operating characteristics be assessed for proposed data combination procedures, including Bayesian approaches, and especially those involving multiple testing. For example, the FDA Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials [20] indicates that demonstration of operating characteristics of Bayesian decision criterion from a frequentist perspective (i.e., expected Type I and Type II error rates) is an essential element in the evaluation of the validity of Bayesian inference for regulatory studies. While these error rates simply measure the frequency of "an assertion of something that is absent", the requirement's implication is that the procedure, when used to determine whether the drug is efficacious over infinitely many samples, must arrive at a wrong conclusion less often on average than some prespecified desired rate. That said, an approved drug must also remain efficacious, with respect to its intended indication, even beyond the patients studied in the trial.

Although this viewpoint is at odds with the Bayesian approach to hypothesis testing, which eschews p-values and reaches decisions solely via posterior probabilities of models or hypotheses

(see e.g. [13]), evaluation the frequentist operating characteristics of a procedure remains good statistical practice. What this implies is that when a drug is deemed significantly better than placebo or significantly better than a threshold value with respect to another drug, it must mean averaging over groups of patients with similar variability over time. A corresponding definition of the Type I error rate here is the probability of observing study (pediatric) data and prior (adult) data that together falsely reject the null hypothesis, as averaged over the variability in both the unknown parameters and the as-yet unobserved data. For example, if the Bayesian decision rule is to reject the null hypothesis when $P(\lambda > 0|D, \boldsymbol{D_0}) \geq p^*$, where λ is the treatment difference and p^* is a prespecified threshold value, we seek a model and prior such that the rate of this decision is less than some upper bound $\alpha \in (0, 1)$, maximized over the variability of unknown parameters and as-yet unobserved data D, when there is indeed no difference.

As has been observed in the discussion of the methods of extrapolation, the use of conjugate pairs can lead to reduced computational burden in evaluating Type I error. In more general cases, this evaluation proceeds by generating a sufficient number of artificial data sets, estimating the mean effect for each data set, and computing the proportion of times that λ has $\mathbb{P}(\lambda > 0) > p^*$, $p^* = 1 - c/2$, for some threshold $1 - c/2$. If the threshold is 0.975 then $c = 0.05$, although it is best to calibrate the threshold value $1 - c/2$ so that the desired Type I error is achieved. But one also has to consider the number of MCMC samples needed to produce a stable estimate of the mean posterior probability over the artificial datasets. For example, suppose there are N data sets generated based on the scenarios that could potentially generate the as-yet unobserved primary data D. For each generated data D_r, the target parameter λ is assumed to have a mean $\mu_\lambda = 0$ and variance σ^2_λ so that the null hypothesis is $H_0 : \lambda \leq 0$ against the alternative $H_1 : \lambda > 0$. Then the Type I error can be approximated as $N^{-1} \sum_{r=1}^{N} I(\widehat{\mathbb{P}}(\lambda > 0|D_r) > 1 - c/2)$, where $\widehat{\mathbb{P}}(\lambda > 0|D_r) = M^{-1} \sum_{r=m}^{M} I(\lambda^{(m)} > 0|D_r)$, with $\lambda^{(m)}$ as the posterior draw for λ after the mth MCMC iteration, $m = 1, \ldots, M$.

3 Examples

Examples of previous use of Bayesian statistics in pediatric drug approvals, at least as reported in the published literature, are limited. However in this section, we summarize previous applications, and also illustrate a subset of the methods outlined in the previous section in pediatric approval settings that can benefit from a Bayesian approach.

3.1 Approval of Remicade® for Pediatric Use

Remicade® (infliximab) is a drug approved by the FDA to treat Crohn's disease, ulcerative colitis (UC), rheumatoid arthritis, and other conditions. Its maker sought a meeting with a gastrointestinal (GI) FDA advisory panel for the purpose of expanding the drug's labeling to include *pediatric* ulcerative colitis. As is common in such settings, extrapolation from adult data was not permitted for dosing or safety assessment in children, but the panel was in favor of the sponsor's argument for extrapolation of efficacy using two existing adult studies. However, no quantitative modeling was involved in this decision; only clinical judgment.

The summary statistics of the data used in the advisory panel can be found in literature (see [55, 56]). The data comprise $K = 2$ UC trials in adults (called ACT 1 and ACT 2) and one UC trial in pediatrics (called T72). The efficacy endpoints in these trials are based on

	ACT 1 (adult)	ACT 2 (adult)	T72 (pediatric)
	Infliximab 5mg/kg	Infliximab 5mg/kg	Infliximab 5mg/kg
Endpoint	$n = 121$	$n = 121$	$n = 60$
Clinical response	84 (69.4%)	78 (64.5%)	44 (73.3%)
Clinical remission	47 (38.8%)	41 (33.9%)	24 (40.0%)
Mucosal healing	75 (62.0%)	73 (60.3%)	41 (68.3%)

Table 2: Study-level endpoint data, Remicade UC studies in adults (ACT 1 and ACT 2) and pediatrics (T72)

the Mayo score derived from the subscores of its 4 components: stool frequency, rectal bleeding, endoscopic findings, and physician's global assessment. Each of these components have subscores 0,1,2, and 3; hence, the Mayo score has a minimum value of 0 and a maximum value of 12. The primary efficacy endpoint of clinical response at week 8 is defined as a decrease in the Mayo score by at least 30% and 3 points, with a decrease in the rectal bleeding subscore of at least 1 point or a rectal bleeding subscore of 0 or 1. When this definition is met, the clinical response is 1; otherwise, it is 0. The secondary endpoints are presence or absence of clinical remission and mucosal healing at week 8. Table 2 gives the summary statistics of the endpoints. In the remainder of this section, we apply Bayesian methods to these data, obtaining quantitative summaries that might have helped the panel make a better-informed decision.

3.1.1 Two-step approach

Let Y_{jk} denote the binary outcome on the primary endpoint for patient j in study k, and $Y_k = \sum_{j=1}^{J} Y_{jk}$ denote the summary statistic for the primary endpoint total in study k^{th}, $k = 1, 2, 3$. Thus, the adult data is $\boldsymbol{D}_0 = (Y_1, Y_2)$ and the pediatric data is $D = Y_3$. We assume that $Y_k \sim \text{Binomial}(n_k, \theta_k)$ and use a conjugate $\text{Beta}(\kappa_\alpha\mu_\alpha, \kappa_\alpha(1-\mu_\alpha))$ prior on θ_k. We complete the model specification by assigning hyperpriors $\kappa_\alpha \sim \text{Uniform}(2, 122)$ and $\mu_\alpha \sim \text{Beta}(1, 1)$, where the upper bound for κ_α was chosen to be comparable to the sample size of one of our adult datasets. As seen in Section 2.2.2, the joint distribution for κ_α and μ_α is $\text{Beta}(\kappa_\alpha\mu_\alpha + \sum_{k=1}^{2} Y_k, \kappa_\alpha(1-\mu_\alpha) + \sum_{k=1}^{2}(n_k - Y_k))$. Note that the support of $(\kappa_\alpha, \mu_\alpha)$ is $(2, 122) \times (0, 1)$. Denote the posterior means of κ_α and μ_α by $\hat{\kappa}_\alpha$ and $\hat{\mu}_\alpha$, and use these in the prior for the pediatric dataset. Specifically, we assume $\theta_3 \sim \text{Beta}(r\hat{\kappa}_\alpha\hat{\mu}_\alpha, r\hat{\kappa}_\alpha(1-\hat{\mu}_\alpha))$ where we use $r \in (0, 1)$ to scale down the adult data effective sample size, $\hat{\kappa_\alpha}$, as the pediatric population (n_3=60) is much smaller than the combined adult populations ($n_1 + n_2 = 242$). Note that r can be either assumed known or have a $\text{Beta}(1, 1)$ prior, and thus in the latter case it is determined from the data. Here, we assumed that r is known. Following similar algebraic routes, we obtain a $\text{Beta}(r\hat{\kappa}_\alpha\hat{\mu}_\alpha + Y_3, r\hat{\kappa}_\alpha(1-\hat{\mu}_\alpha) + n_3 - Y_3)$ posterior for θ_3. Note that the empirical pediatric success rate is slightly higher than that of the adults, and hence when we borrow more strength from the adult data, the θ_3 estimates should decrease, and the corresponding 95% credible intervals should get narrower.

The results for various choices of r (corresponding to ESS values ranging from 62.4 up to the full combined sample size of 302) are given in the first two columns of Table 3. Notice

Two Step Prior		Commensurate Prior		Power Prior	
r (ESS)	$E(\theta_3\|D, \boldsymbol{D}_0)$ (CrI)	κ_a (ESS)	$E(\theta_3\|D, \boldsymbol{D}_0)$ (CrI)	a_0(ESS)	$E(\theta_3\|D, \boldsymbol{D}_0)$ (CrI)
0.01 (62.4)	0.729 (0.616,0.832)	1 (62.4)	0.730 (0.622,0.837)	0 (62)	0.725 (0.615, 0.835)
0.25 (121)	0.687 (0.607,0.761)	10 (84)	0.724 (0.620,0.828)	0.25 (123)	0.710 (0.631, 0.790)
0.5 (181)	0.667 (0.602,0.728)	50 (181)	0.705 (0.617,0.793)	0.5 (183)	0.705 (0.638, 0.770)
1 (302)	0.659 (0.606,0.711)	100 (302)	0.700 (0.622,0.778)	1 (304)	0.700 (0.648, 0.752)

Table 3: Results for the various Bayesian models fit to the study-level Remicade data.

that as we increase the r, the θ_3 point estimate decreases (from 0.729 down to 0.659) and the corresponding 95% credible interval widths also decrease (from 0.216 down to 0.105), both as expected. In this model, the choice of r is somewhat subjective, but values greater that 0.5 result in $(181 - 60)/60 \approx 201\%$ or more of θ_3 posterior's strength coming from the adult data, which may be excessive.

3.1.2 Combined approaches

In this section we change notation slightly, assuming $Y_k \sim \text{Binomial}(n_k, \theta_0)$ for adult study $k = 1, 2$, and $Y_3 \sim \text{Binomial}(n_3, \theta_3)$ for the pediatric population. We first fit a commensurate prior model. In particular, we choose a $\text{Beta}(1, 1)$ distribution as our initial prior on θ_0 and take $\theta_3|\theta_0 \sim \text{Beta}(\kappa\theta_0, \kappa(1 - \theta_0))$ as our commensurate prior with $\kappa \sim \text{Gamma}(\kappa_\alpha, 1)$ and κ_α assigned a fixed value. Following equation (8), the joint posterior in this case arises as

$$\pi(\theta_3, \theta_0, \kappa|D, \boldsymbol{D}_0) \propto \theta_3^{Y_3}(1-\theta_3)^{n_3-Y_3}\theta_0^{Y_1+Y_2}(1-\theta_0)^{n_1+n_2-Y_1-Y_2}\theta_3^{\kappa\theta_0-1}(1-\theta_3)^{\kappa(1-\theta_0)-1}\kappa^{\kappa_\alpha-1}e^{-\kappa}.$$

While this does not lead to a closed form for the marginal posterior $p(\theta_3|D, \boldsymbol{D}_0)$, sampling from the distribution is routine via the `BUGS` language (see code in Appendix A). The results for varying values of κ_α are given in Table 3; the ESS values shown for this method are computed as functions of our posterior estimates $\widehat{\kappa}_\alpha$. As the with the two-step approach, increases in κ_α are again associated with clear decreases in the θ_3 point estimates and interval widths, though the shrinkage back to the adult values is less dramatic here.

Finally, we also apply the power prior method to our data, also as described in Section 2.2.2. We fix the powers $a_{0,k} = a_0$ for $k = 1, 2$, thus specifying a fixed and equal amount of borrowing from both adult studies. Notationally, we have $Y_k \sim \text{Binomial}(n_k, \theta_3)$ and $\theta_3 \sim \text{Beta}(\kappa\mu, \kappa(1 - \mu))$, where we choose non-informative values $\kappa = 2$ and $\mu = 0.5$. Then $\theta_3|D, \boldsymbol{D}_0 \sim \text{Beta}(\tilde{\kappa}\tilde{\mu}, \tilde{\kappa}(1-\tilde{\mu}))$, where $\tilde{\kappa}$ and $\tilde{\mu}$ as given in Section 2.2.2. The results for four representative values of α_0 (and corresponding ESS values) are given in the last column of Table 3. The effect of increasing ESS is again apparent, with the now-familiar trends in the posterior means and interval widths being somewhat intermediary to those arising from the previous two methods.

In all of these approaches, borrowing information from adults led to more precise calculation of the treatment response in pediatrics. In trial design settings, this can mean meaningful sample size reductions. We remark that the use of individual patient-level data from these three studies may enable even better estimates of θ_3, especially if accompanied by corresponding information on explanatory covariates (say, patient age, gender, baseline health status, etc.).

3.2 Extrapolating Adult Exposure-Response Information of Anti-epileptic Drugs to Pediatrics

Trileptal® (oxcarbazepine) is an anti-epileptic drug (AED) approved by FDA in adults and pediatrics as adjunctive therapy and as monotherapy for the treatment of partial seizures. The pediatric monotherapy indication was obtained by pharmacokinetic bridging from adult monotherapy trials, based on similarity in the exposure-response (PK/PD) relationship between adults and pediatrics in the adjunctive therapy setting. The similarity of the PK/PD relationship in adults and pediatrics in the adjunctive therapy was assessed quantitatively using frequentist approaches in the clinical pharmacology FDA review document [57]. The bridging approach leveraged information from adults and resulted in the pediatric monotherapy indication without the need for a clinical trial in pediatrics. Recently, the FDA released a policy statement [58] that extrapolation of efficacy from adults to pediatric patients 4 years of age and older for adjunctive therapy of partial seizures is acceptable based on quantitative PK and PK/PD assessments of several AEDs approved so far. Assessing the similarity of the PK/PD relationship between adults and pediatrics by borrowing information from adults for extrapolation in pediatrics fits naturally into the Bayesian paradigm.

In this subsection, the application of Bayesian approaches, discussed in Section 2, is demonstrated for assessing similarity in the exposure-response relationship on a simulated exposure-response data in adults and pediatrics based on a Trileptal clinical pharmacology review. The bridging approach which leverages information from adults is illustrated as well. The exposure-response data was generated using the dose-exposure (C_{min}: minimum plasma concentration) relationship and the exposure-response (Percent change from baseline (PCB) in 28-day seizure frequency) relationship available in the review document. The dose-C_{min} relationship for both adults and pediatrics is as follows:

$$\log(C_{min}) = \eta_0 + \eta_1 \times (\text{isAED}) + \eta_2 \times \log(\text{dose in } mg/m^2/\text{day}) + \epsilon, \tag{11}$$

where $\epsilon \sim \text{N}(0, \sigma^2)$. The dose was derived based on generating typical adult and pediatric age (4–18 years), body weight and body surface distributions, and isAED is an indicator variable for the presence of interacting AEDs. The simulations assumed 1:1 allocation of placebo and treatment in both adults and pediatrics.

The C_{min}-PCB in 28-day seizure frequency data in adults and pediatrics was generated based on simulated individual C_{min} and assuming the following model:

$$\log(\text{PCB} + 110) = \zeta_0 + \zeta_1 \times C_{min} + \zeta_2 \times C_{min} \times (\log(\text{baseline seizure frequency}) - 2.5) + \varepsilon, \tag{12}$$

where $\varepsilon \sim \text{N}(0, \tau^2)$. In this model, one concludes that the exposure-response relationship is the same between adults and pediatrics if the two groups have the same value for the slope, ζ_1. An interaction effect between C_{min} and baseline seizure frequency is also included. Overall, individual C_{min} and PCB seizure frequency data was simulated for 464 adults and 221 pediatrics, similar to the review. The simulated mean C_{min} in adults and pediatrics were 86 μmol/L and 84 μmol/L respectively, leading to a mean percent change from baseline in seizure frequency of -61%.

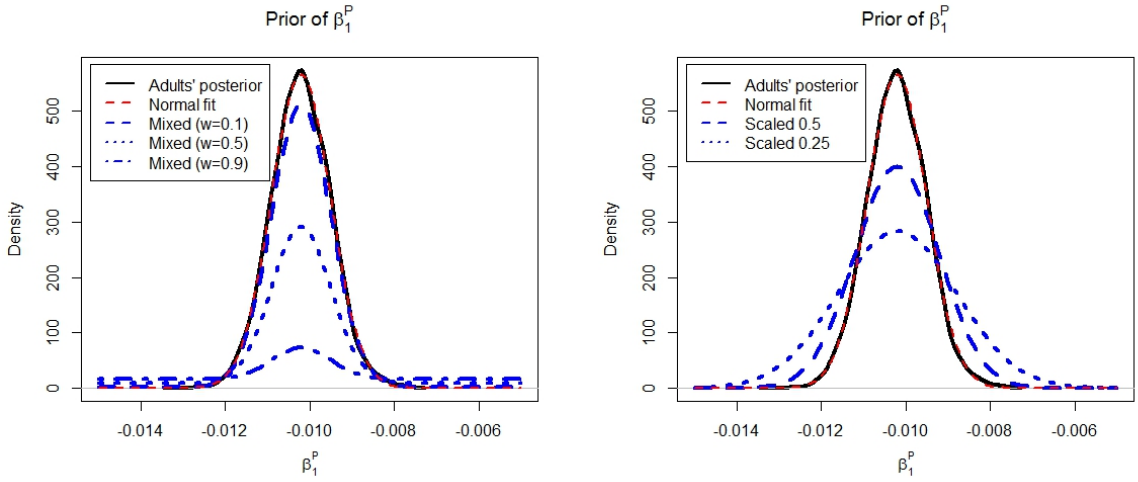
Parameter	Posterior Mean (SD)	Median (95% Cred Int)
α_0	4.55 (0.04)	4.55 (4.47, 4.63)
α_1	-0.0102 (0.0007)	-0.0102 (-0.0116, -0.0088)
α_2	0.0035 (0.0007)	0.0035 (0.0022, 0.0049)
τ_α	0.652 (0.022)	0.651 (0.611, 0.637)
Diagnostics		
DIC	923	
Mean deviance	919	

Table 4: Posterior distribution summaries from model for adults (including intercept term)

3.2.1 Two-step approach

Let α_0, α_1, α_2, and τ_α^2 be the parameters of (12) that is used to fit the adult data, $\boldsymbol{D}_0$. These parameters are given vague priors $\alpha_0 \sim \mathrm{N}(0, 100)$, $\alpha_1 \sim \mathrm{N}(0, 100)$, $\alpha_2 \sim \mathrm{N}(0, 100)$, and $\tau_\alpha^2 \sim \mathrm{IG}(0.001, 0.001)$. Through routine MCMC sampling, their posterior means and associated credible intervals are given in Table 4 and we determined that the posterior density $\hat{p}(\alpha_1|\text{adult})$ is best represented by a normal density with mean -0.0102 and variance that is empirically calculated from the MCMC samples. This is then used to form the prior for the slope β_1 in the model that is used to fit the pediatric data. Its form follows the robust mixture (5), i.e., $\hat{\pi}(\beta_1) = w_I \hat{p}(\alpha_1|D_0) + w_F \mathrm{N}(0, 1000)$, where $w_I = 1 - w_F$ and the priors for the intercept β_0, baseline seizure frequency adjustment β_2, and sampling variance τ_β^2 are left vague like the adult model. The shape of this prior is given in Figure 1(a) where w_I is given varying weights to the informative and flat component, e.g., $w_I = 1$ means complete borrowing while $w_I = 0$ means no borrowing. Table 5 shows the results of the estimate of the posterior means and the density of the MCMC samples for β_1 are shown in Figure 2(a) again under varying weights to the informative and flat component. The posterior probability that β_1 is in the 95% credible interval of α_1, without borrowing information from adults, is 69.3%. This implies that we would not rule out that the two slopes are the same if the threshold probability that the two slopes are different is set at 2.5%. This result also points to some consistency between adult and pediatric populations as the posterior distribution falls in the contours of the posterior distribution of the treatment response parameter in the adult data (cf. [59]). Furthermore, the results also point that full borrowing appears to fit the pediatric data best.

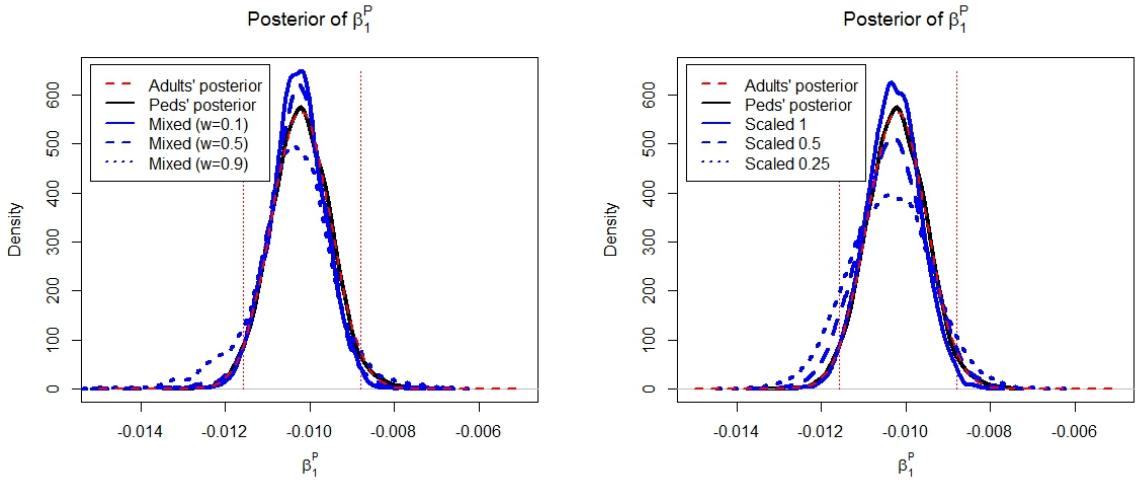
We also investigated a different way of making the prior for β_1 robust to prior-data conflict by variance scaling (see Figure 1(b)). In variance scaling, the sample variance $Var(\alpha_1)$ that is empirically calculated from the MCMC samples for α is re-scaled by a factor ϖ when it is used as a prior for the slope β_1 in the model that is used to fit the pediatric data, i.e., $\pi(\beta_1) = \mathrm{N}(\hat{\alpha}_1, Var(\alpha_1)/\varpi)$, where $\hat{\alpha}_1$ is the estimate of α. When $0 < \varpi < 1$, this corresponds to diffusing the prior centered at $\hat{\alpha}_1$. Table 6 gives the results for the estimates of the posterior mean and Figure 2(b) gives the posterior density of b_1 at varying degrees of scaling. This table also gives the posterior probability that β_1 is in the 95% credible interval of α_1. One expects that as $\varpi \to 0$ this posterior probability should approach 69.3%.



(a) Robust mixture

(b) Scaled variance

Figure 1: Robustification of the normal approximation to the adults' posterior distribution of β_1 in the Two-Step approach by (a) mixing with a flat distribution into a mixture prior as given by a fixed weight w (w = weight given to the flat component), or (b) scaling of the variance by a factor k.



(a) Robust mixture

(b) Scaled variance

Figure 2: Combined Approach: (a) Posterior distribution (kernel density) using robust-mixture priors (w = weight given to the flat component of the mixture). (b) Posterior distribution (kernel density) using variance-scaled priors.

3.2.2 Combined approaches

We also apply the power prior method, described in Section 2.2.2, using fixed power a_0. Let $m_\beta(i) = \beta_0 + \beta_1 \times C_{min} + \beta_2 \times C_{min} \times (\log(\text{baseline seizure frequency}) - 2.5)$ be the model fitted to the pediatric data and let $m_\alpha(i) = \alpha_0 + \beta_1 \times C_{min} + \alpha_2 \times C_{min} \times (\log(\text{baseline seizure frequency}) - 2.5)$ the model fitted for the adult data. Note that the two models have the same parameter for the slope β_1. Suppose $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau_\alpha^2, \tau_\beta^2)$ denotes the prior distribution for the parameters $\boldsymbol{\alpha} = (\alpha_0, \alpha_2)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$. Specifically, we assume that α_0 and β_0 are exchangeable $\mathrm{N}(\mu_0, \tau_0^2)$, α_2 and β_2 are exchangeable $\mathrm{N}(\mu_2, \tau_2^2)$, and $\mu_0, \tau_0^2, \mu_2, \tau_2^2$ are vague. Then because α_0 is fixed, the joint posterior is proportional to

$$\begin{aligned} p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau_\alpha^2, \tau_\beta^2 | \boldsymbol{D}_0, D) \quad \propto \quad & \prod_{i=1}^{N_{\text{peds}}} \phi(m_\beta(i), \tau_{\text{peds}}^2 | D) \left[\prod_{j=1}^{N_{\text{adult}}} \phi(m_\alpha(j), \tau_{\text{adult}}^2 | \boldsymbol{D}_0) \right]^{a_0} \\ & \times \quad \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau_\alpha^2, \tau_\beta^2) \pi(\mu_0, \tau_0^2, \mu_2, \tau_2^2), \end{aligned} \tag{13}$$

where $\phi(c, d)$ is the normal density function with mean c and variance d. Here borrowing primarily happens for the parameter β_1 as it is common to the likelihood of the pediatrics and adults but there could also be borrowing happening with the other parameters because of the exchangeability assumption. Estimation proceeds, once again, using MCMC and Table 7 shows the results for different values of a_0. As expected, as $a_0 \to 1$ the credible interval of β_1 is narrowest and the borrowed information is 1024 which is more than the sample size of the adults. One possible reason is that the definition of ESS as suggested by [46] may not be accurate for the parameters in the linear model and can be explored more in future research.

Lastly, we applied a generalization of the commensurate prior approach to the adult and pediatric data. Let $m_\beta(i) = \beta_0 + \beta_1 \times C_{min} + \beta_2 \times C_{min} \times (\log(\text{baseline seizure frequency}) - 2.5)$ be the model fitted to the pediatric data and let $m_\alpha(i) = \alpha_0 + \alpha_1 \times C_{min} + \alpha_2 \times C_{min} \times (\log(\text{baseline seizure frequency}) - 2.5)$ the model fitted for the adult data. Here, the model fitted to the pediatric and adult data have distinct parameters. To induce borrowing, we center the prior of β_1 on α_1, i.e., $\beta_1 | \alpha_1, \iota_1 \sim \left[\mathrm{N}(\alpha_1, \tau^{-1})\right]^{1-\iota_1} \left[\mathrm{N}(\alpha_1, R^{-1})\right]^{\iota_1}$, $\iota_1 \sim \text{Bernoulli}(\nu_1)$ and $\tau \sim \text{Uniform}(s_l, s_u)$. We specify the spike so that if $\beta_1 | \alpha_1 \sim \mathrm{N}(\alpha_1, R^{-1})$, β_1 deviates negligibly from α_1. On the other hand, we specify the slab to contain small values that correspond to modest shrinkage of β_1 to α_1. The tails of the prior distribution are controlled by s_l and s_u where when $s_l \approx 0$, smaller values of s_u provide a prior with heavier tails. We specify R to be large relative to the magnitude of β_1 and α_1, e.g., $R = 2000$, and is ν_k chosen so that the DIC is the smallest. For β_0 and β_2, we used a very similar specification but instead of fixing the value of ν_0 and ν_2, we used $\text{Beta}(1, 1)$ (see [43] for more details on generalized commensurate priors). Lastly, the rest of the parameters $(\alpha_0, \alpha_1, \alpha_2, \tau_\alpha^2, \tau_\beta^2)$ are again given vague priors. Table 7 shows the results which shows modest borrowing of information between α_1 and β_1 which may be viewed positively in regulatory settings. While many values of ν_1 were evaluated, the value of $\nu_1 = 0.75$ was chosen to minimize deviance.

To summarize the points made in this example, we used Bayesian approaches to get the posterior probability of exposure in the two groups to determine if they are different. While the statistical evaluation did not rule out that they are different, much of this evaluation relies on clinical/pharmacological meaning. Then the bridging approach was applied to leverage information from adults into a more precise estimation of exposure response in pediatrics.

4 Discussion and Conclusions

The overviews of the current use of Bayesian methods in pediatric research presented above contribute to understanding where more guidance might be needed. Increased use of one or a combination of these approaches might significantly reduce operational efforts related to the execution of pediatric trials where the population is limited, or the disease itself is considered as a rare disease. It is acknowledged that execution of these steps should follow proper planning, and every effort should be made to avoid any bias or false positive conclusions about the treatment. After initial efforts to create the above-mentioned overviews, there is a clear need to continue to examine aspects of the Bayesian approach in order to suggest its optimized use in pediatric research, especially for more advanced clinical stages (e.g. Phase III).

4.1 Optimized Bayesian approach to pediatric research

The scope of this paper includes medicinal products for which efficacy and safety has been shown in adults and further development in children is to be conducted. This assumes that further use in children is to be submitted as an extension of an already-approved indication in adults. As such, we rely on existing data and their optimal use in designing the pediatric studies and providing medicines to children as soon as possible without impacting the medicinal product's safety, efficacy and quality. Because of this inherent borrowing of information about therapeutic response in adults to pediatrics, the Bayesian approach (which embeds priors information into the estimation of response parameters in the current data) fits pediatric drug development in a natural way. Certainly, the extrapolation has to be dependent on the physiological plausibility as outlined in many extrapolation guidances (see [8, 60, 21]). The approach should reduce sample size needed, assuming congruence of pediatric data with adult data, without sacrificing the ability to estimate effect with reasonable precision. This should also translate into shorter trial durations for the pediatric population.

The discussion of common practical obstacles with using Bayesian in pediatrics and ideas how they can be resolved is ongoing. As mentioned in Section 1.2, the machinery to create the design and perform the analysis of pediatric trials is limited. This, as well as a paucity of Bayesian education for statisticians and key decision makers, deters Bayesian methodology from being more broadly applied. These obstacles manifest themselves into issues like "malleability" of methods, lack of standardization, and opaque processes. Bayesian methods are sometimes incorrectly viewed as alternative routes to approval for medicinal products for which current data evidence is weak. Optimizing the use of Bayesian methodology in pediatric research requires a transparent and streamlined operation in both the design and analysis of the trial. In particular, methods and assumptions need to be laid bare, estimation diagnostics and model fit thoroughly explored, priors need to be robust to the effects of prior-data conflict, consistency or compatibility of target (pediatric) and reference (adult) data checked, sensitivity or results investigated, prior effective sample size determined, and operational characteristics (bias, type I error rate and power) evaluated.

As mandated by regulations, plans for testing every new medicinal product intended for pediatric use should be prospectively agreed with authorities in the context of a Pediatric Study Plan (PSP) in the USA, or a Paediatric Investigation Plan (PIP) in Europe. We encourage that all aspects of prospectively planned Bayesian statistics intended to be used in pediatric trials be

discussed with regulators through these plans. These discussions may take into consideration prior information, its weight, comparability of studies from which the prior information would derive, covariates, etc.

4.2 Timely Regulatory Interactions

Pediatric research usually faces recruitment difficulties caused by low disease incidence in children, operational difficulties, or general reluctance to include children in clinical trials [11]. Independent of the cause, pediatric trials are becoming more global, and sponsors are searching for adequate clinical research sites in Europe, the USA, and other countries. Because of this, the main aim is to have trials harmonized as much as possible, which is a current challenge. The use of more complex statistical methods will increase the level of interaction among authorities, but since these discussions are currently done separately with each authority, the chances for discrepancies are greater.

Since 2009, a parallel scientific advice procedure with the EMA and FDA can be requested by applicants. Specifically, in the last 4 years (2011-2014) a total of 14 parallel scientific advice requests were made, compared with 1496 requests for EMA scientific advice made in the same time span (i.e., less than 1% of all requests) [61]. The parallel scientific advice might be beneficial to learn the positions of both agencies at the same time, but the problem of further harmonization remains [62]. Moreover, the advice provided is not "joint" advice, but only "parallel"; both agencies reserve their rights to assess provided data separately (i.e., their "final answers" might be still different). It ought to be noted that parallel scientific advice is reserved for products of special interest to both agencies, and there is no guarantee that the advice will be granted, acknowledging both agencies' resource constraints. Both agencies seem to have commitment to develop a framework for facilitating pediatric research, in terms of facilitating regular exchange of scientific and ethical issues and other information on pediatric development programmes in Europe and the US to avoid exposing children to unnecessary trials, as well as developing global pediatric development plans based on scientific grounds, and compatible for both agencies [63].

A more advanced collaboration action has been seen with release of the joint EMA/FDA proposal for research in Gaucher disease in 2014, where more detailed proposals for possible extrapolations and study designs were outlined [64]. This is seen as a helpful guide to applicants. On the other hand, it is difficult to expect to have such an approach in every single condition acknowledging that the duration of preparation of the joint proposal for Gaucher disease took almost 3 years.

In summary, we believe that the Bayesian approach can be utilized towards the increased use of already available information on product's safety, efficacy, and quality (SEQ) to inform pediatric development. This might be incorporated seamlessly through early dialogues with agencies (EMA and FDA) through scientific advice and pediatric plans which could include considerations on possibilities for conditional approval of new pediatric indications based on the amount of previously available information on product's SEQ and increased use of post-approval risk management tools and commitments.

Acknowledgment

We would like to thank Clara Dominguez-Islas, PhD for helping us in some of the computations in this manuscript.

References

[1] R English, Y Lebovitz, and R Griffin. Institute of Medicine (US) Forum on Drug Discovery Development and Translation. Transforming clinical research in the United States. Challenges and opportunities: workshop summary, 2010.

[2] US Congress. Best pharmaceuticals for children act. `http://www.fda.gov/opacom/laws/pharmkids/pharmkids.html`, 2002.

[3] US Congress. Pediatric research equity act. `http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108_cong_public_laws&docid=f:publ155.108`, 2003.

[4] Sherry G Selevan, Carole A Kimmel, and Pauline Mendola. Identifying critical windows of exposure for children's health. *Environmental Health Perspectives*, 108(Suppl 3):451, 2000.

[5] Mark D Miller, Melanie A Marty, Amy Arcus, Joseph Brown, David Morry, and Martha Sandy. Differences between children and adults: implications for risk assessment at california epa. *International Journal of Toxicology*, 21(5):403–418, 2002.

[6] Michelle Roth-Cline, Jason Gerson, Patricia Bright, Catherine S Lee, and Robert M Nelson. Ethical considerations in conducting pediatric research. In *Pediatric Clinical Pharmacology*, pages 219–244. Springer, 2011.

[7] Patrina HY Caldwell, Sharon B Murphy, Phyllis N Butow, and Jonathan C Craig. Clinical trials in children. *The Lancet*, 364(9436):803–811, 2004.

[8] European Medicines Agency. Concept paper on extrapolation of efficacy and safety in medicine development. `http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/04/WC500142358.pdf`, 2013.

[9] European Medicines Agency. Reflection paper on extrapolation of efficacy and safety in pediatric medicine development. `http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2016/04/WC500204187.pdf`, 2016.

[10] Yeruk Mulugeta, Jeffrey S Barrett, Robert Nelson, Abel Tilahun Eshete, Alvina Mushtaq, Lynne Yao, Nicole Glasgow, Andrew E Mulberg, Daniel Gonzalez, Dionna Green, et al. Exposure matching for extrapolation of efficacy in pediatric drug development. *The Journal of Clinical Pharmacology*, 2016.

[11] Julia Dunne, William J Rodriguez, M Dianne Murphy, B Nhi Beasley, Gilbert J Burckart, Jane D Filie, Linda L Lewis, Hari C Sachs, Philip H Sheridan, Peter Starke, et al. Extrapolation of adult data and other data in pediatric drug-development programs. *Pediatrics*, 128(5):e1242–e1249, 2011.

[12] Hari Sachs. Pediatric extrapolation. http://www.fda.gov/downloads/NewsEvents/MeetingsConferencesWorkshops/UCM415474.pdf, 2014.

[13] Bradley P Carlin and Thomas A Louis. *Bayesian Methods for Data Analysis, 3rd edition*. Boca Raton, FL: CRC Press, 2009.

[14] Scott M Berry, Bradley P Carlin, J Jack Lee, and Peter Muller. *Bayesian Adaptive Methods for Clinical Trials*. Boca Raton, FL: CRC Press, 2010.

[15] Steven N Goodman and John T Sladky. A Bayesian approach to randomized controlled trials in children utilizing information from adults: the case of Guillain-Barre. *Clinical Trials*, 2(4):305–310, 2005.

[16] David A Schoenfeld, Hui Zheng, and Dianne M Finkelstein. Bayesian design using adult data to augment pediatric trials. *Clinical Trials*, 6(4):297–304, 2009.

[17] Marc Gastonguay. Application of population pharmacokinetic and pharmacodynamic modeling and simulation n. http://www2.cde.org.tw/action/uploadfile/2012/0327/GastonguayTaiwanNatlUExamples.pdf, 2009.

[18] Ian Wadsworth, Lisa Victoria Hampson, and Thomas Friedrich Jaki. Extrapolation of efficacy and other data to support the development of new medicines for children: a systematic review of methods. *Statistical Methods in Medical Research*, 2016 (to appear).

[19] John T Mondick, Leonid Gibiansky, Mark R Gastonguay, Jeffrey M Skolnik, Michael Cole, Garett J Veal, Alan V Boddy, Peter C Adamson, and Jeffrey S Barrett. Population pharmacokinetic investigation of actinomycin-d in children and young adults. *The Journal of Clinical Pharmacology*, 48(1):35–42, 2008.

[20] U.S. Department of Health and Human Services. Guidance for the use of Bayesian statistics in medical device clinical trials. www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071072.htm, 2010.

[21] U.S. Department of Health and Human Services. Leveraging existing clinical data for extrapolation to pediatric uses of medical devices. http://www.fda.gov/ucm/groups/fdagov-public/@fdagov-meddev-gen/documents/document/ucm444591.pdf, 2015.

[22] Laura Thompson and Gene Pennello. Borrowing from adult data to make inferences about the effectiveness of medical devices in a pediatric population. *Technical Report, Food and Drug Administration, Center for Devices and Radiological Health*, 2015.

[23] CMPH. Guideline on clinical trials in small populations, 2006.

[24] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53, 1763.

[25] Allan Birnbaum. On the foundations of statistical inference. *Journal of the American Statistical Association*, 57(298):269–306, 1962.

[26] David J Spiegelhalter, Nicola G Best, Bradley P Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

[27] David Lunn, Chris Jackson, Nicky Best, Andrew Thomas, and David J. Spiegelhalter. *The BUGS book: A practical introduction to Bayesian analysis*. Boca Raton, FL: CRC Press, 2012.

[28] José M Bernardo. The concept of exchangeability and its applications. *Far East Journal of Mathematical Sciences*, 4:111–122, 1996.

[29] Saman Muthukumarana and Ram C Tiwari. Meta-analysis using Dirichlet process. *Statistical Methods in Medical Research*, page 0962280212453891, 2012.

[30] Heinz Schmidli, Sandro Gsteiger, Satrajit Roychoudhury, Anthony O'Hagan, David Spiegelhalter, and Beat Neuenschwander. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70(4):1023–1032, 2014.

[31] Mike West. Approximating posterior distributions by mixture. *Journal of the Royal Statistical Society, Ser. B*, 55:409–422, 1993.

[32] Jose M. Bernardo and Adrian F.M. Smith. *Bayesian Theory*. New York: Wiley, 1994.

[33] Thomas S Ferguson. Prior distributions on spaces of probability measures. *The Annals of Statistics*, pages 615–629, 1974.

[34] Anthony O'Hagan, Luis Pericchi, et al. Bayesian heavy-tailed models and conflict resolution: A review. *Brazilian Journal of Probability and Statistics*, 26(4):372–401, 2012.

[35] Joseph G Ibrahim and Ming-Hui Chen. Power prior distributions for regression models. *Statistical Science*, 15:46–60, 2000.

[36] Meg A Gamalo, Ram C Tiwari, and Lisa M LaVange. Bayesian approach to the design and analysis of non-inferiority trials for anti-infective products. *Pharmaceutical Statistics*, 13(1):25–40, 2014.

[37] Yuyan Duan, Keying Ye, and Eric P Smith. Evaluating water quality using power priors to incorporate historical information. *Environmetrics*, 17(1):95–106, 2006.

[38] Hong Zhao, Brian P. Hobbs, Haijun Ma, Qi Jiang, and Bradley P. Carlin. Combining non-randomized and randomized data in clinical trials using commensurate priors. *Research report, University of Minnesota Division of Biostatistics*, 2015.

[39] Brian P Hobbs, Bradley P Carlin, Sumithra J Mandrekar, and Daniel J Sargent. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, 67(3):1047–1056, 2011.

[40] Brian P Hobbs, Daniel J Sargent, and Bradley P Carlin. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis*, 7(3):639–674, 2012.

[41] Jairo A Fúquene, John D Cook, Luis R Pericchi, et al. A case for robust Bayesian priors with applications to clinical trials. *Bayesian Analysis*, 4(4):817–846, 2009.

[42] Thomas A. Murray, Brian P. Hobbs, Theodore C. Lystig, and Bradley P. Carlin. Semiparametric bayesian commensurate survival model for post-market medical device surveillance with non-exchangeable historical data. *Biometrics*, 70:185–191, 2014.

[43] Thomas A. Murray, Brian P. Hobbs, and Bradley P. Carlin. Combining nonexchangeable functional or survival data sources in oncology using generalized mixture commensurate priors. *Annals of Applied Statistics*, 9:1549–1570, 2015.

[44] David J. Spiegelhalter, Keith R. Abrams, and Jonathan P. Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. New York: Wiley, 2004.

[45] Satoshi Morita, Peter F Thall, and Peter Müller. Determining the effective sample size of a parametric prior. *Biometrics*, 64(2):595–602, 2008.

[46] Donald Malec. A closer look at combining data among a small number of binomial experiments. *Statistics in Medicine*, 20(12):1811–1824, 2001.

[47] Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–472, 1992.

[48] Adrian E Raftery and Steven M Lewis. Comment: One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science*, 7:493–497, 1992.

[49] Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. Coda: Convergence diagnosis and output analysis for MCMC. *R News*, 6:7–11, 2006.

[50] Stan Development Team. Stan: A C++ library for probability and sampling, version 2.8.0, 2015. URL `http://mc-stan.org/`.

[51] Martyn Plummer. Jags: Just another Gibbs sampler, 2004. URL `http://mcmc-jags.sourceforge.net/`.

[52] Alan E Gelfand, Dipak K Dey, and Hong Chang. Model determination using predictive distributions with implementation via sampling-based methods (with discussion). *Bayesian Statistics 4*, pages 147–167, 1992.

[53] Alan E. Gelfand and Dipak K. Dey. Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Ser. B*, 56:501–514, 1994.

[54] European Parliament and of the Council of 12 December 2006. Regulation (EC) No 1901/2006 of the European Parliament and of the Council of 12 December 2006. `http://ec.europa.eu/health/files/eudralex/vol-1/reg_2006_1901/reg_2006_1901_en.pdff`, 2016.

[55] Jeffrey Hyams, Lakshmi Damaraju, Marion Blank, Jewel Johanns, Cynthia Guzzo, Harland S Winter, Subra Kugathasan, Stanley Cohen, James Markowitz, Johanna C Escher, et al. Induction and maintenance therapy with infliximab for children with moderate to severe ulcerative colitis. *Clinical Gastroenterology and Hepatology*, 10(4):391–399, 2012.

[56] Jeffrey S Hyams, Trudy Lerer, Anne Griffiths, Marian Pfefferkorn, Michael Stephens, Jonathan Evans, Anthony Otley, Ryan Carvalho, David Mack, Athos Bousvaros, et al. Outcome following infliximab therapy in children with ulcerative colitis. *The American Journal of Gastroenterology*, 105(6):1430–1436, 2010.

[57] U.S. Department of Health and Human Services. NDA 21014/S003 Pharmacology and Biopharmaceutical Review(s). `http://www.accessdata.fda.gov/drugsatfda_docs/nda/2003/021014_S003_TRILEPTAL$\%2$0TABLETS_BIOPHARMR.pdf`, 2003.

[58] Pediatric News. 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). *Journal of Pediatric Pharmacology and Therapeutics*, 21(1):98–98, 2016.

[59] A Lawrence Gould, Tian Jin, Li Xin Zhang, and William WB Wang. A predictive bayesian approach to the design and analysis of bridging studies. *Journal of Biopharmaceutical Statistics*, 22(5):916–934, 2012.

[60] U.S. Department of Health and Human Services. General Clinical Pharmacology Considerations for Pediatric Studies for Drugs and Biological Products. `http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm425885.pdf`, 2014.

[61] European Medicines Agency. EMA Annual Report 2014. `http://www.ema.europa.eu/docs/en_GB/document_library/Annual_report/2015/04/WC500186306.pdf`, 2014.

[62] European Medicines Agency and U.S. Food and Drug Administration . General Principles, EMA-FDA Parallel Scientific Advice. `http://www.ema.europa.eu/docs/en_GB/document_library/Other/2009/11/WC500014868.pdf`, 2009.

[63] European Medicines Agency and U.S. Food and Drug Administration. Principles of Interactions: Between EMEA and FDA Pediatric Therapeutics. `http://www.ema.europa.eu/docs/en_GB/document_library/Other/2009/12/WC500017969.pdf`, 2007.

[64] European Medicines Agency . Gaucher disease: A strategic collaborative approach from EMA and FDA. `http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2014/05/WC500166587.pdf`, 2014.

Appendix A

`BUGS` code for the commensurate prior model described in Section 3.1.2:

```
model{
theta3~dbeta(k*theta0,k*(1-theta0)) #The commensurate prior
theta0~dbeta(alpha0,beta0) #The initial prior

alpha0<-1
beta0<-1

k~dgamma(kalpha,1)

Y[1]~dbinom(theta0,n[1])
Y[2]~dbinom(theta0,n[2])
Y[3]~dbinom(theta3,n[3])

}
```

w_l	w_F	Parameter	Posterior Mean (SD)	Posterior Median (95% Cred Int)	Mean Deviance	DIC	Posterior Probability*	ESS†
		β_0	4.57 (0.06)	4.57 (4.46, 4.68)				
1	0	β_1	-0.0103 (0.0006)	-0.0103 (-0.0115, -0.0090)	482.5	484.1	97.2	1005
		β_2	0.0046 (0.0010)	0.0046 (0.0025, 0.0066)				
		τ_β	0.733 (0.035)	0.731 (0.669, 0.805)				
		β_0	4.57 (0.06)	4.57 (4.46, 4.68)				
0.9	0.1	β_1	-0.0103 (0.0006)	-0.0103 (-0.0115, -0.0090)	485.0	487.6	97.0	962
		β_2	0.0046 (0.0010)	0.0046 (0.0025, 0.0066)				
		τ_β	0.733 (0.034)	0.732 (0.669, 0.804)				
		β_0	4.57 (0.06)	4.57 (4.45, 4.68)				
0.5	0.5	β_1	-0.0103 (0.0007)	-0.0103 (-0.0117, -0.0090)	485.7	489.3	95.4	801
		β_2	0.0046 (0.0010)	0.0046 (0.0025, 0.0066)				
		τ_β	0.733 (0.034)	0.732 (0.669, 0.804)				
		β_0	4.57 (0.06)	4.57 (4.45, 4.68)				
0.1	0.9	β_1	-0.0104 (0.0010)	-0.0103 (-0.0125, -0.0085)	484.4	488.0	87.0	428
		β_2	0.0046 (0.0011)	0.0046 (0.0025, 0.0068)				
		τ_β	0.733 (0.034)	0.732 (0.669, 0.805)				
		β_0	4.57 (0.06)	4.57 (4.45, 4.70)				
0	1	β_1	-0.0105 (0.0013)	-0.0105 (-0.0131, -0.0079)	484.9	489	69.3	221
		β_2	0.0047 (0.0013)	0.0047 (0.0022, 0.0072)				
		τ_β	0.734 (0.034)	0.733 (0.670, 0.805)				

Table 5: Posterior distribution summaries from model for paediatrics, using robust-mixture priors (w_F = weight given to the flat component of the mixture, $w_I = 1 - w_F$ = weight given to the informative component of the mixture [the posterior from the adult data]). *Posterior probability of β_1 (pediatric) within the 95% CI of β_1 (adults), expressed as percentage. †Using the posterior variance of b_1 using a flat prior $w_F = 1$ as numerator.

ϖ	Parameter	Posterior Mean (SD)	Posterior Median (95% Cred Int)	Mean Deviance	DIC	Posterior Probability*	ESS†
0.5	β_0	4.57 (0.06)	4.57 (4.45, 4.68)	485.7	488.9	91.2	606
	β_1	-0.0103 (0.0008)	-0.0103 (-0.0119, -0.0088)				
	β_2	0.0046 (0.0011)	0.0046 (0.0024, 0.0067)				
	τ_β	0.733 (0.034)	0.731 (0.671, 0.804)				
0.25	β_0	4.57 (0.06)	4.57 (4.46, 4.69)	485.8	490.6	83.7	412
	β_1	-0.0104 (0.0010)	-0.0104 (-0.0123, -0.0084)				
	β_2	0.0046 (0.0011)	0.0046 (0.0024, 0.0069)				
	τ_β	0.734 (0.035)	0.732 (0.668, 0.806)				

Table 6: Posterior distribution summaries from model for pediatrics, using variance-scaled priors. *Posterior probability of β_1 (pediatric) within the 95% CrI of α_1 (adults), expressed as percentage. †Using the posterior variance of the posterior distribution of β_1 using a flat prior $w_F = 1$ as numerator.

α_0	Parameter	Posterior Mean (SD)	Posterior Median (95% Cred Int)	Mean Deviance	DIC	Posterior Probability*	ESS†
Power Prior							
1	β_0	4.57 (0.06)	4.57 (4.45, 4.68)	486.5	489.7	97.4	1024
	β_1	-0.0104 (0.0006)	-0.0103 (-0.0115, -0.0091)				
	β_2	0.0046 (0.0010)	0.0046 (0.0026, 0.0065)				
	τ_β	0.733 (0.034)	0.732 (0.668, 0.802)				
0.5	β_0	4.57 (0.06)	4.57 (4.46, 4.68)	484.8	488.2	91.0	592
	β_1	-0.0103 (0.0008)	-0.0105 (-0.0119, -0.0088)				
	β_2	0.0046 (0.0011)	0.0047 (0.0025, 0.0067)				
	τ_β	0.733 (0.035)	0.731 (0.670, 0.804)				
0.25	β_0	4.57 (0.06)	4.57 (4.45, 4.69)	484.0	487.2	83.5	404
	β_1	-0.0104 (0.0011)	-0.0105 (-0.0123, -0.0085)				
	β_2	0.0046 (0.0011)	0.0046 (0.0024, 0.0068)				
	τ_β	0.733 (0.035)	0.731 (0.668, 0.803)				
0	β_0	4.57 (0.06)	4.57 (4.45, 4.70)	485.4	492.1	68.5	221
	β_1	-0.0105 (0.0013)	-0.0106 (-0.0131, -0.0079)				
	β_2	0.0047 (0.0013)	0.0047 (0.0023, 0.0072)				
	τ_β	0.733 (0.034)	0.732 (0.671, 0.803)				
Commensurate Prior							
0.75	β_0	4.56 (0.04)	4.56 (4.48, 4.63)	1401	1408	75.1	270
	β_1	-0.0104 (0.0012)	-0.0104 (-0.0127, -0.0080)				
	β_2	0.0047 (0.0013)	0.0047 (0.0022, 0.0071)				
	τ_β	0.727 (0.035)	0.726 (0.661, 0.797)				

Table 7: Posterior distribution summaries from model for pediatrics using power priors. *Posterior probability of β_1 (pediatric) within the 95% CrI of α_1 (adults), expressed as percentage.†Using the posterior variance of β_1 using a flat prior as numerator.