Citation: CPT Pharmacometrics Syst. Pharmacol. (2016) 00, 00; doi:10.1002/psp4.12092

ORIGINAL ARTICLE

Model-Based Assessment of Alternative Study Designs in Pediatric Trials. Part II: Bayesian Approaches

G Smania[1,2], P Baiardi[3], A Ceci[1], M Cella[1,4]* and P Magni[2]

This study presents a pharmacokinetic-pharmacodynamic based clinical trial simulation framework for evaluating the performance of a fixed-sample Bayesian design (BD) and two alternative Bayesian sequential designs (BSDs) (i.e., a non-hierarchical (NON-H) and a semi-hierarchical (SEMI-H) one). Prior information was elicited from adult trials and weighted based on the expected similarity of response to treatment between the pediatric and adult populations. Study designs were evaluated in terms of: type I and II errors, sample size per arm (SS), trial duration (TD), and estimate precision. No substantial differences were observed between NON-H and SEMI-H. BSDs require, on average, smaller SS and TD compared to the BD, which, on the other hand, guarantees higher estimate precision. When large differences between children and adults are expected, BSDs can return very large SS. Bayesian approaches appear to outperform their frequentist counterparts in the design of pediatric trials even when little weight is given to prior information from adults.
CPT Pharmacometrics Syst. Pharmacol. (2016) **00**, 00; doi:10.1002/psp4.12092; published online on 0 Month 2016.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ Bayesian approaches are appealing in pediatric trials because the required sample size can be significantly reduced by taking prior information into account. When the efficacy of a compound to be investigated in the pediatric population has already been established in clinical studies in older populations, evidence available from such studies can be used as "prior."

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ This study presents a PK-PD based clinical trial simulation framework able to evaluate the performance of three BDs in pediatric trials using prior information from adults' data.

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE
☑ Benefits from the use of BDs are maximized when children respond similarly to adults, although advantages compared to standard approaches are still present in case of differences between the two populations.

HOW THIS MIGHT CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS
☑ PK-PD CTS of Bayesian study designs in children has never been performed before: this approach allows investigating the influence of a pediatric study design on trial success before exposing actual children to the treatment.

In the sibling paper accompanying this article,[1] we introduced the need for alternative study designs in the implementation of randomized controlled trials (RCTs) for the evaluation of efficacy and safety of an experimental treatment in the pediatric population and we presented frequentist approaches of alternative designs.

In this second part of the work, we focus on study designs for pediatric trials based on Bayesian approaches. Bayes' theorem allows one to make inference on observed data by incorporating *a priori* beliefs (usually defined in terms of a prior probability distribution) on the phenomenon being observed. From an RCT perspective, historical information on treatment effect (e.g., from previous studies) can be leveraged to infer the efficacy of the treatment being studied in the new RCT.[2] Consequently, compared to classical frequentist approaches, the amount of data to be collected in the new study is reduced because these data are augmented by historical ones. This ultimately allows reducing the sample size of the study.

Such property of Bayesian designs (BDs) is of tremendous importance for pediatric trials, where the number of patients that can be recruited is often very limited. In particular, if the disease being studied in the pediatric population is similar to the corresponding disease in an older population, available RCTs in the latter can be leveraged to elicit a prior distribution for treatment effect to be used in the analysis of the pediatric trial (e.g., adult data as prior in pediatric trials, adolescents data as prior for trials in children, children data as prior for trials in infants, etc.). Bayesian techniques are also endorsed by the European Medicine Agency for the use of the extrapolation approach in pediatric drug development programs.[3]

Nonetheless, few examples can be found in the literature on the application of Bayesian approaches in pediatric RCTs borrowing prior information from adults' data.[4–6] Among BDs, those of sequential nature are of further potential interest because of their inherent flexibility, especially if compared with their frequentist counterparts.[7] Although limited, applications of Bayesian sequential

[1]Consorzio per Valutazioni Biologiche e Farmacologiche, Via Luigi Porta 14, Pavia, Italy; [2]Dipartimento di Ingegneria Industriale e dell'Informazione, Università degli Studi di Pavia, Via Ferrata 5, Pavia, Italy; [3]Direzione Scientifica Centrale, Fondazione Salvatore Maugeri, IRCCS, Via Salvatore Maugeri 4, Pavia, Italy; [4]Department of Clinical Pharmacology, Global Clinical Development, Chiesi Farmaceutici, Parma, Italy. *Correspondence: M Cella (mcella@cvbf.net)
Received 26 October 2015; accepted 3 June 2016; published online on 0 Month 2016. doi:10.1002/psp4.12092

Table 1 Investigated scenarios for the evaluation of the performance of the Bayesian design and of the two Bayesian sequential designs.

		BSD	
	BD	**NON-H**	**SEMI-H**
Scenario 1	$v = 0.18$	$n_T = 16.5$	$v = 0.18$
		$n_P = 16.5$	$\omega = 33$
		$p_s = 0.99$	$p_s = 0.99$
		$p_f = 0.5$	$p_f = 0.5$
Scenario 2	$v = 0.4$	$n_T = 3.5$	$v = 0.4$
		$n_P = 3.5$	$\omega = 7$
		$p_s = 0.99$	$p_s = 0.99$
		$p_f = 0.75$	$p_f = 0.75$

BD, Bayesian design; BSD, Bayesian sequential design; NON-H, non-hierarchical; SEMI-H, semi-hierarchical.

designs (BSDs) spanned from early phase II anticancer trials,[8,9] safety monitoring,[10,11] and dose-finding studies,[12] whereas applications to pediatric efficacy trials remain scant.

The scope of the present article is to compare the performance of a standard BD with that of two BSDs by means of pharmacokinetic-pharmacodynamic (PK-PD) based clinical trial simulation (CTS). To the best of our knowledge, no attempts were made in the evaluation of BSDs through PK-PD CTS. Designs are evaluated in terms of: type I and type II errors; sample size per arm (SS); total trial duration (TD); and precision of treatment effect estimate. Finally, results of these BDs are cross-compared with those of the frequentist designs presented in ref. 1.

METHODS

For information on the PK-PD model used to simulate data, the general features of the study design (treatment groups, doses, primary endpoints, and trial duration) and the framework for CTS the reader should refer to our companion article.[1] Only methods differentiating from such an article are reported thereafter.

In agreement with the Bayesian approach, the null hypothesis of no treatment difference (H_0) was tested through the posterior probability of the improvement provided by topiramate (TPM; the drug under study as additional therapy to the current patient specific antiepileptic treatment) over placebo (in addition to the current patient specific antiepileptic treatment) in epileptic children (δ_P) after having observed the clinical trial data (i.e., $p(\delta_P|Data)$).

Study designs description

Bayesian design. In a two arm BD, patients are randomized to two parallel groups to receive either placebo or TPM, with the number of patients to be randomized in each group fixed *a priori*.

The statistical framework for the BD was adapted from Schoenfeld *et al.*[5] Formally, let δ_A and δ_P be the true improvements of TPM over placebo in the adult and pediatric population, respectively. Their prior distribution is $\delta_A, \delta_P \sim N(\mu, v^2)$, where μ has a non-informative prior ($\mu \sim N\left(0, \sigma_\mu^2\right)$, $\sigma_\mu^2 \to +\infty$), while v is given a fixed number reflecting the supposed difference in improvement of TPM over placebo between children and adults. The parameter v plays a pivotal role in the design and analysis of the trial. Schoenfeld *et al.*[5] suggest eliciting v from clinical judgment or, if available, from previous pediatric and adult studies as $v = |\bar{\delta}_A - \bar{\delta}_P|/\sqrt{2}$, where $\bar{\delta}_A$ and $\bar{\delta}_P$ represent estimates of δ_A and δ_P obtained from historical data. Because in the PK-PD model used for CTS[13] both pediatric and adult data were modelled (**Supplementary Table S1**), we deduced plausible values of v from parameter estimates of the final PK-PD models in the two populations. In particular, Monte Carlo methods were used to obtained $\bar{\delta}_P$ from 10^6 samples, which was set equal to 0.2467 as in our companion article,[1] whereas $\bar{\delta}_A$ (set to 0.5016) was obtained in the same way of $\bar{\delta}_P$ but using adult PK-PD parameters and an average adult TPM dose regimen of 150 mg b.i.d.[14,15] This led us to set $v = 0.18$ (hereafter called scenario 1). In order to explore different scenarios and to take into account the plausible situation of a larger difference in TPM improvement over placebo between children and adults, v was set to 0.4 (hereafter called scenario 2, see **Table 1**).

Furthermore, let $\hat{\delta}_A$ be the maximum likelihood estimate of TPM effect over placebo based on an adult trial with $m_A/2$ patients per arm, and $\hat{\delta}_P$ the same estimate in the new pediatric trial with $m_P/2$ patients per arm. We focused on the case of $\hat{\delta}_A$ and $\hat{\delta}_P$ normally distributed, in particular $\hat{\delta}_A \sim N\left(\delta_A, {s_A^2}/{m_A}\right)$ and $\hat{\delta}_P \sim N\left(\delta_P, {s_P^2}/{m_P}\right)$, with $s_A = 2\sigma_A$ and $s_P = 2\sigma_P$ (σ_A and σ_P are the SDs of Y in the adult and pediatric population, respectively). We assumed $\sigma_A = \sigma_P = \sigma$, with the value of σ given by the PK-PD model and reported in **Supplementary Table S1**. Quantitatively, adult prior information was incorporated by setting $\hat{\delta}_A$ to the model-derived value of $\bar{\delta}_A$ and m_A to 663, which corresponds to the number of adult patients used to identify the adult PK-PD model.

According to the statistical framework from Schoenfeld *et al.*,[5] because δ_P, $\hat{\delta}_P$, and $\hat{\delta}_A$ follow a multivariate normal distribution with mean 0 and covariance matrix:

$$\begin{bmatrix} \sigma_\mu^2 + v^2 & \sigma_\mu^2 + v^2 & \sigma_\mu^2 \\ & \sigma_\mu^2 + v^2 + {s_P^2}/{m_P} & \sigma_\mu^2 \\ & & \sigma_\mu^2 + v^2 + {s_A^2}/{m_A} \end{bmatrix},$$

the posterior distribution of δ_P is its conditional distribution given $\hat{\delta}_P$ and $\hat{\delta}_A$. By letting $\sigma_\mu^2 \to +\infty$ it turns out that $\delta_P|\hat{\delta}_P, \hat{\delta}_A \sim N\left(\mu_{\delta_P}, {\sigma_{\delta_P}}^2\right)$ where

$$\mu_{\delta_P} = \frac{\frac{m_P}{S_P^2}\hat{\delta}_P + \frac{\omega}{S_A^2}\hat{\delta}_A}{\frac{m_P}{S_P^2} + \frac{\omega}{S_A^2}} \quad (1)$$

$$\sigma_{\delta_P}{}^2 = \frac{s_P^2 s_A^2}{m_P s_A^2 + \omega s_P^2}, \quad (2)$$

with $\omega = \frac{m_A s_A^2}{s_A^2 + 2v^2 m_A}$. μ_{δ_P} depicts the Bayesian estimator of the improvement of TPM over placebo in the pediatric population.

In their work, Schoenfeld *et al.*[5] provide a method to define a Bayesian analogue of classical frequentist power given by the following formula:

$$Bayesian\_Power\left(\delta_P^*\right) = \Phi\left(\frac{\sqrt{m_P}}{s_P}\left[\delta_P^* - \frac{s_P{}^2}{m_P}\left(z_{1-\alpha}\sqrt{\frac{m_P}{s_P{}^2} + \frac{\omega}{s_A{}^2}} - \frac{\omega}{s_A{}^2}\hat{\delta}_A\right)\right]\right) \quad (3)$$

with z_x being the x-th quantile of the standard normal distribution, δ_P^* the minimum clinically important difference in TPM vs. placebo in the pediatric population, and Φ the normal cumulative distribution function. If $v \to \infty$ (no data are borrowed from adults), $\omega \to 0$ and Eq. 3 collapses to the classical frequentist power.

Similarly to the approach used in the frequentist setting,[1] the sample size of the study was identified by exploiting Eq. 3, fixing Bayesian power to 0.8, α to 0.05, and δ_P^* to $\bar{\delta}_P$, that is, to 0.2467. According to the calculated total sample size at each iteration of step 3 of the CTS framework (see ref. 1) half of the patients were assigned to the placebo group and half to the TPM group. Step 4 consisted in H_0 acceptance/rejection based on the posterior probability of treatment effect: if $p\left(\delta_P > 0 | \hat{\delta}_P, \hat{\delta}_A\right) \leq 1-\alpha$ H_0 is accepted, otherwise it is rejected.

BSDs: non-hierarchical and semi-hierarchical framework

In a sequential design, statistical analyses are sequentially performed after the enrollment of groups of patients of predetermined size *G*. This allows early stopping of the trial for either efficacy or futility. In our companion article,[1] we considered two alternative implementations of frequentist sequential designs. In the present work, we consider two Bayesian implementations of sequential designs adapted from Gsponer *et al.*[16]: one in a non-hierarchical (NON-H) and one in a semi-hierarchical (SEMI-H) framework. The NON-H is presented in **Supplementary Material S1**, while the SEMI-H is detailed hereafter.

The SEMI-H shares the same framework of the BD, but, because it is a sequential design, inferences from the posterior distribution of δ_P are sequentially made at each interim analysis. Accordingly, $\hat{\delta}_P$ and s_P are computed at each interim analysis rather than being estimated once at the end of the trial. The decisional criteria for trial success/failure (i.e., rejection/acceptance of H_0) used for BSDs were the following:

$$\begin{cases} Success\ (H_0\ rejected): \ p\left(\delta_P > 0 | \hat{\delta}_P, \hat{\delta}_A\right) > p_s \\ Failure\ (H_0\ accepted): \ p\left(\delta_P < \delta_{\min} | \hat{\delta}_P, \hat{\delta}_A\right) > p_f \end{cases} \quad (4)$$

where $\delta_{\min}$ was set to 0.12, which corresponds to a 10% further decrease in seizure reduction for TPM 7 mg/kg/day against placebo, considering an average placebo seizure reduction of 21.5% (obtained from the PK-PD model). The parameters of the posterior distribution used to evaluate criteria (4) in the SEMI-H correspond to that of BD (Eqs. 1 and 2). Consistently with the BD, NON-H and SEMI-H performance have been investigated under two alternative scenarios in terms of v, p_s, and p_f (**Table 1**).

BSDs were simulated with $G = 20$. In agreement with the sequential nature of these designs, for each CTS, steps 3 and 4 of the procedure described in ref. 1 were sequentially performed until trial success/failure was detected according to criteria (4). In particular, in step 3, *G* children were randomized to TPM and placebo in a 1:1 ratio and their simulated responses used to compute the posterior probabilities in (4) based on the NON-H and the SEMI-H, while in step 4 such probabilities were compared with their corresponding thresholds.

Supplementary Material S1 contains a detailed description of the calculation of the metrics used for design comparison for the BD, NON-H, and SEMI-H. The R code for simulation of BD, NON-H, and SEMI-H under scenario 1 can be found in **Supplementary Material S2**.

RESULTS

In this section, only the results obtained under scenario 1 are presented, whereas results from scenario 2 are reported in **Supplementary Material S1**.

Type I and type II errors

Because of the inherently different philosophy of the Bayesian approach compared to the frequentist one, there is no control of type I error in the design of Bayesian trials and its value depends upon the weight of prior information. In our case, because prior information came from a successful adult trial, type I errors resulted in 22.2% for the BD, 26.3% for the NON-H, and 24.1% for the SEMI-H. As expected, type II errors for the BD is around its predetermined value of 20% (**Table 2**). In BSDs, probably due to the increased type I errors, it is slightly smaller: 15.1% for the NON-H and 16.0% for the SEMI-H, despite it was not planned at the design stage.

Sample size per arm

SS for the BD is determined *a priori* on the basis of Eq. 3 and its relationship with v is shown in **Figure 1a**. Unlike BD, SS of BSDs is not known *a priori* and the histograms of SS achieved at each simulation in the two scenarios are depicted in **Figure 1b**.

For all designs, the average SS required in scenario 1 is lower than that of scenario 2 because, in the latter, less weight is given to prior information from adult data. In the BD, the required SS resulted in 49 children per arm, which approximately corresponds to the 75th percentile of SS distributions of both BSDs (**Table 2**), distributions that do not significantly differ between each other (**Figure 1b**).

Average SS for the NON-H and SEMI-H is 37 (**Table 2**). Despite mean and median SS of BSDs are lower than the BD one, sequential recruitment of children may be remarkably prolonged if treatment effect signals are captured later

Table 2 Performance metrics obtained from 1,000 clinical trial simulations of the investigated designs

		BD		BSD			
				NON-H		SEMI-H	
Design metric		**Scenario 1**	**Scenario 2**	**Scenario 1**	**Scenario 2**	**Scenario 1**	**Scenario 2**
$\hat{\alpha}$ (%) (95% CI)		22.2 (19.6–24.8)	7.3 (5.7–8.9)	26.3 (22.6–29.0)	9.4 (7.6–11.2)	24.1 (21.4–26.8)	7 (5.4–8.6)
$\hat{\beta}$ (%) (95% CI)		20.3 (17.8–22.8)	21.1 (18.6–23.6)	15.1 (12.9–17.3)	19.5 (17.0–22.0)	16.0 (13.7–18.3)	19.4 (16.9–21.9)
E(SS) (patients)		49	103	37	67	37	66
SS_{50} (patients)		49	103	20	50	20	50
SS_{75} (patients)		49	103	50	90	50	90
SS_{90} (patients)		49	103	81	150	80	150
SS_{95} (patients)		49	103	110	190	110	190
Median TD	ER = 4 patients/months	26.5	53.5	12	27	12	27
	ER = 10 patients/months	11.8	22.6	6	12	6	12

$\hat{\alpha}$, type I error; $\hat{\beta}$, type II error; SS_{50}, median sample size per arm; SS_{75}, SS_{90}, SS_{95}, 75th, 90th, and 95th percentiles of sample size distribution); BD, Bayesian design; BSD, Bayesian sequential design; CI, confidence interval; ER, enrollment rate; NON-H, non-hierarchical; SEMI-H, semi-hierarchical; SS, sample size; TD, total trial duration.

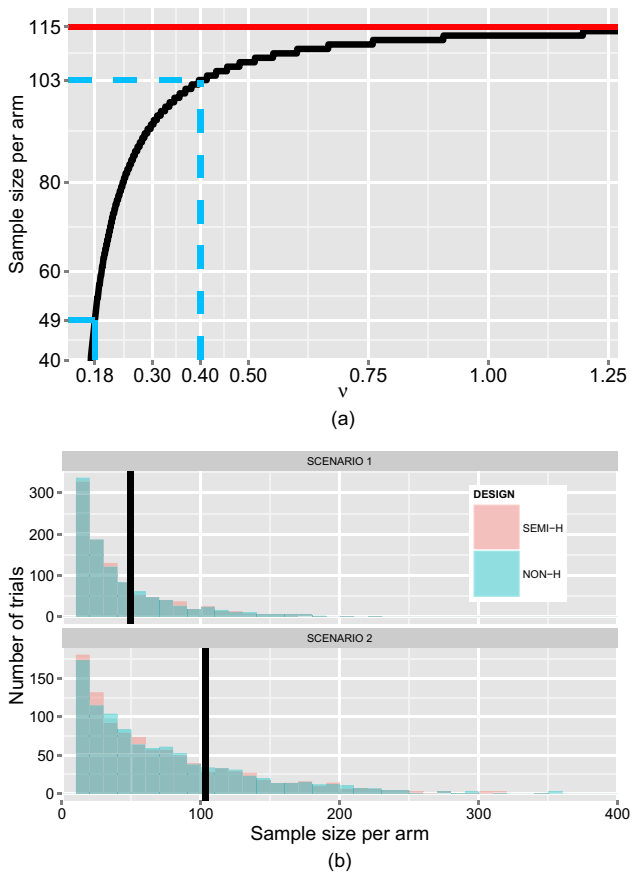

Figure 1 (**a**) Sample size per arm (SS) vs. difference in improvement of topiramate (TPM) over placebo between children (δ_P) and adults (δ_A) expressed in terms of SD of the prior distribution on δ_P and δ_A (ν). The red line represents the sample size per arm of a classical parallel frequentist design, azure lines indicate the value of ν and the corresponding SS of the Bayesian design (BD) in scenario 1 (solid line) and 2 (dotted line). (**b**) Histograms of SSs obtained at each of the 1,000 clinical trial simulation of the Bayesian sequential design in the non-hierarchical (NON-H; green histogram) and semi-hierarchical (SEMI-H; pink histogram) framework for scenario 1 (upper panel) and 2 (lower panel). The black vertical lines indicate SS of the BD in the two scenarios.

on during the trial, as shown by the right tail of the histograms in upper panel of **Figure 1b**. The probability of the NON-H and SEMI-H requiring a higher SS than the BD is about 27%.

Total trial duration

TD as a function of enrollment rate is shown in **Figure 2**. The NON-H and SEMI-H have the lowest median duration among the three designs, which reflects the lowest median SS required (**Table 2**). Likely due to the lack of significant difference between SS distributions in NON-H and SEMI-H, their medians and 95% prediction intervals in TD perfectly overlap.

Treatment difference estimate (μ_{δ_P}) precision

Figure 3a shows that the BD leads on average to the shortest width of the 95% credible intervals thereby guaranteeing the highest precision in the Bayesian estimate of δ_P in both evaluated scenarios. When less weight is given to prior information on adult (right panel of **Figure 3a**), the precision appears to increase in all investigated designs.

No significant differences can be detected between the precision assured by NON-H and SEMI-H in both scenarios.

DISCUSSION

Comparison within Bayesian designs

General. PK-PD CTS provides a favorable tool to integrate prior information on drug disposition and effect to evaluate the performance of candidate study designs. In this work, we underpinned CTS with a PK-PD model, which was separately identified from both pediatric and adult data.[13] This further enabled us to formalize CTS of BDs by using prior information from adults' data.

In particular, we were able to quantify the difference in improvement of TPM over placebo between children and adults (ν) starting from parameter estimates of the PK-PD model. The decision on the value of ν is critical because it ultimately affects both the required SS and the final estimate of δ_P and the associated inference. However, the

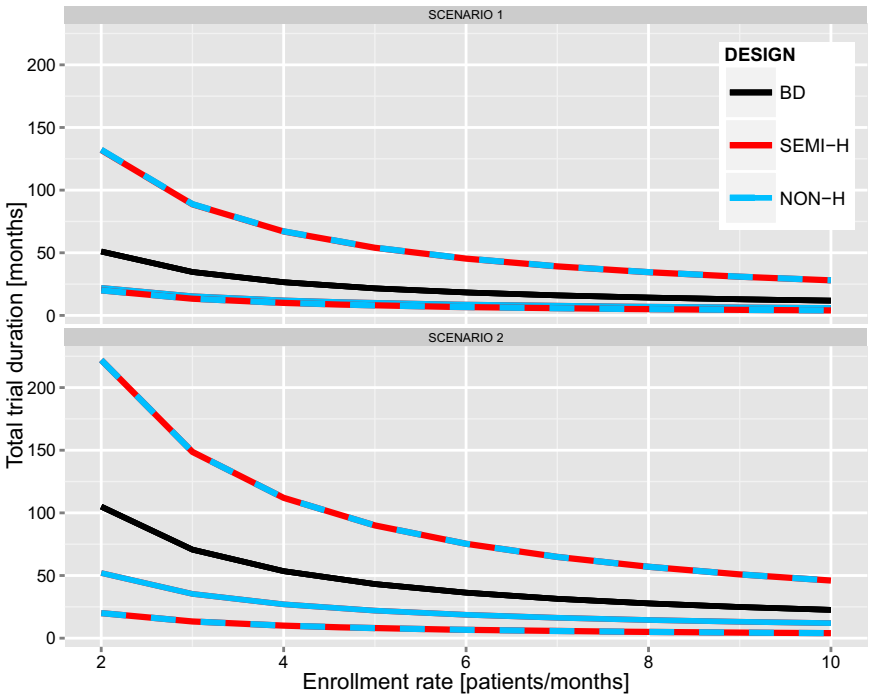

Figure 2 Total trial duration as a function of enrollment rate for the Bayesian design (black line) and the Bayesian sequential design in the non-hierarchical (NON-H; light blue lines) and semi-hierarchical (SEMI-H; orange lines) framework for scenario 1 (upper panel) and 2 (lower panel). Solid lines represent the median duration whereas dotted lines depict 95% prediction intervals. Median and 95% prediction intervals of NON-H and SEMI-H are on top of each other in both scenarios.

value of v depends upon the specific problem and is not universal for all compounds and/or diseases. In our analysis, for $v > 1$ the required SS of the BD tends toward that required by a standard frequentist design (**Figure 1a**). On the other hand, the value of v obtained from PK-PD model parameters (scenario 1) leads to a SS ∼60% lower than that required in a frequentist setting, with clear advantages from a patient's recruitment perspective.

Comparing both investigated scenarios, **Figure 3b** shows that, for all designs, the estimate of δ_P (i.e., μ_{δ_P}, the posterior mean of TPM improvement over placebo in seizures reduction) shifts toward the pediatric value given by the PK-PD model if a greater v is considered. Also, **Figure 3b** suggests that BSDs led to an estimate of δ_P closer to the adult value when compared to their fixed-sample counterpart, partly because of the lower SS required by sequential designs, which makes μ_{δ_P} to rely more on prior (adult) information.

Should the pediatric PK-PD model in the pediatric population not be available at the design stage, a model-based approach still provides considerable benefits in eliciting prior information. If, for example, children are expected to be twice as sensitive as adults (assumption that can be supported for instance by historical data from drugs with similar mechanism of action in children), v can be derived by using the adult PK-PD model with a doubled drug-effect parameter. Consequently, the impact of such an assumption on designs performance can be quantitatively evaluated by means of our framework.

NON-H vs. SEMI-H

Our results show that there are no significant differences between the NON-H and the SEMI-H in both investigated scenarios and across all analyzed metrics (**Table 2**). Such results were not totally unexpected as the weight of prior information in NON-H and SEMI-H is given in an equivalent manner (see **Supplementary Material S1**). Nonetheless, because SEMI-H explicitly enables to weight prior information on adults on the grounds of clinical and scientific plausibility, it may be preferable to NON-H where the weight of prior information is to be assigned based on an "equivalent sample size."

BSDs vs. BD

BSDs, under comparable type I errors, require, on average, a lower SS than the BD (**Table 2**). With respect to mean SS, the reduction seen under scenario 2 (∼35%) is slightly higher than that observed under scenario 1 (∼24%). Although limited by only two scenarios, this suggests that the more the adult and pediatric populations are different the greater is the advantage in terms of mean SS brought by BSD compared to a fixed-sample BD. On the other hand, SS distributions of BSDs become more skewed when passing from scenario 1 to scenario 2 (**Figure 1b**) and more caution is needed because there is a higher probability of ending the trial with non-practical SSs, which would ultimately jeopardize the conclusions that can be drawn from the study.

Differences in TD between BD and BSDs can be explained by the aforementioned differences in SS.

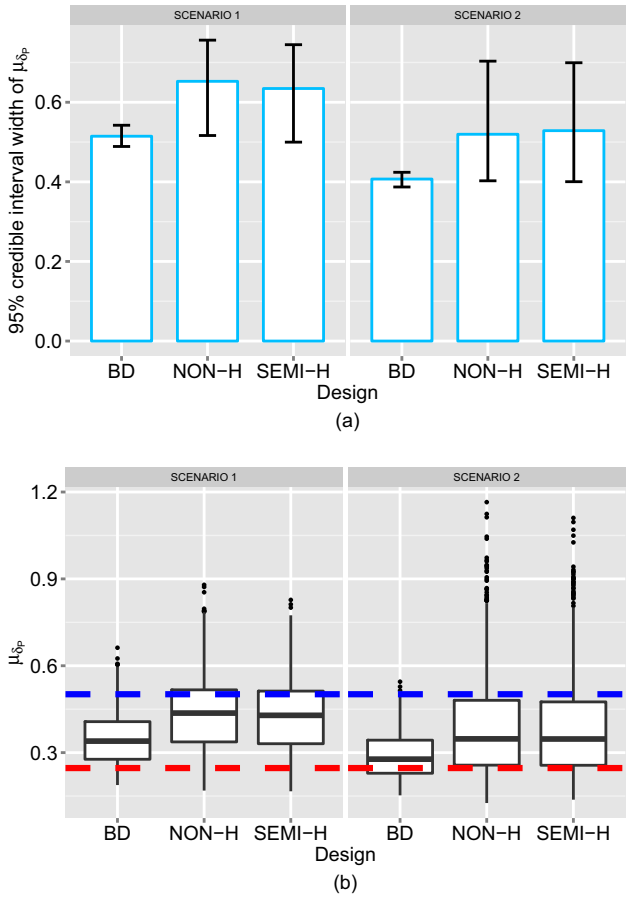

Figure 3 (**a**) Bar chart plot of the median 95% credible interval width of treatment difference estimates (μ_{δ_P}) obtained at each of the 1,000 clinical trial simulation of the Bayesian design (BD), the Bayesian sequential design with a non-hierarchical (NON-H) and a semi-hierarchical (SEMI-H) framework for scenario 1 (left panel) and 2 (right panel). The upper and lower "hinges" correspond to the first and third quartiles of 95% confidence intervals widths. (**b**) Boxplots of μ_{δ_P} obtained at each simulation of the BD, the NON-H and the SEMI-H for scenario 1 (left panel) and 2 (right panel). The blue and red dotted lines represent the adult and pediatric treatment effects difference between topiramate and placebo (obtained from the pharmacokinetic-pharmacodynamic model), respectively.

Figure 2 shows that for the investigated range of enrollment rates median TD of the NON-H and SEMI-H is lower than the duration of the BD. Like for SS, upper limit of 95% prediction interval of TD (upper dotted lines of **Figure 2**) clearly suggests that there exists a low probability of the trial lasting more than 100 months, which would thereby compromise its feasibility.

Regarding the precision of μ_{δ_P}, the BD performs better than the BSDs with a median width of its 95% credible interval about 20% lower than the corresponding median of BSDs (**Figure 3a**). The better precision of the BD is likely due to the higher number of samples used to compute μ_{δ_P}. Moving from scenario 1 (left panel of **Figure 3a**) to scenario 2 (right panel of **Figure 3a**), the median of the 95% credible interval width consistently decreases for all designs, suggesting that, on average, the increase in precision due to a higher SS outweighs the decrease caused by a less informative prior. However, outliers of credible interval widths in BSDs are higher in scenario 2 than in scenario 1 (results not shown). These less precise estimates are obtained when H_0 is rejected at the first interim analysis (i.e., when 10 patients per arm have been enrolled), revealing that for such low SS prior information is not strong enough to guarantee an acceptable precision of μ_{δ_P}.

We also performed a sensitivity analysis in order to investigate the impact of non-negligible model misspecification on BD performance. Results and discussion on such analysis are reported in **Supplementary Material S1**.

Comparison between Bayesian and frequentist approaches

In part I of this work, we presented the performance of a set of alternative frequentist study designs (crossover, randomized withdrawal, sequential probability ratio test (SPRT), and triangular test (TT)) for pediatric trials and compared them with the standard parallel design (PaD).[1] The present article deals with the evaluation of BDs, whose comparison was based on the same metrics used for frequentist designs except for the percentage of exposure to placebo, TPM, and no-treatment relative to total trial exposure, as this measure is equal to that of PaD considering the equal randomization to the two treatment arms.

One of the pivotal issues addressed by our work is the simultaneous comparison of a battery of alternative designs based on a pharmacometric model of the compound and the related placebo effect. Although interesting, comparing the goodness of Bayesian and frequentist approaches is not trivial because of the inherently different philosophy of these two methodologies and is still an open debate.[17] Scenario 1 highlights why BDs are appealing in pediatrics: the required SS is significantly reduced compared to that of their frequentist counterparts (BD vs. PaD (fixed-sample designs) and SEMI-H vs. SPRT/TT (sequential designs)). In particular, SS of BD (49 patients) is nearly 60% lower than that of the PaD (115 patients), while the SS distribution of the SEMI-H is squeezed toward lower SSs compared to that obtained with the SPRT/TT (left panel of **Figure 4**). **Figure 5a** shows that the reduced SS implies a remarkable lower precision of the BD estimate compared to that of the PaD (i.e., $\hat{\delta}$), whereas for BSDs such difference is less pronounced because of the low precisions associated with the SPRT and TT[1]; moreover, the estimated treatment effect is shifted toward the adult value for both fixed-sample (mean μ_{δ_P} of 0.3498) and sequential (mean μ_{δ_P} equals 0.4330 and 0.4265 for the NON-H and SEMI-H, respectively) BDs when compared to the corresponding fixed-sample (mean $\hat{\delta}$ of 0.2821) and sequential (mean $\hat{\delta}$ equals 0.3717 and 0.3588 for the SPRT and TT, respectively) frequentist designs (**Figure 5b**).

Different considerations can be made if frequentist designs are compared with BDs under scenario 2.

In terms of SS, the BD allows to reduce the number of children to be enrolled by almost 10% when compared to the PaD while maintaining similar estimates of treatment

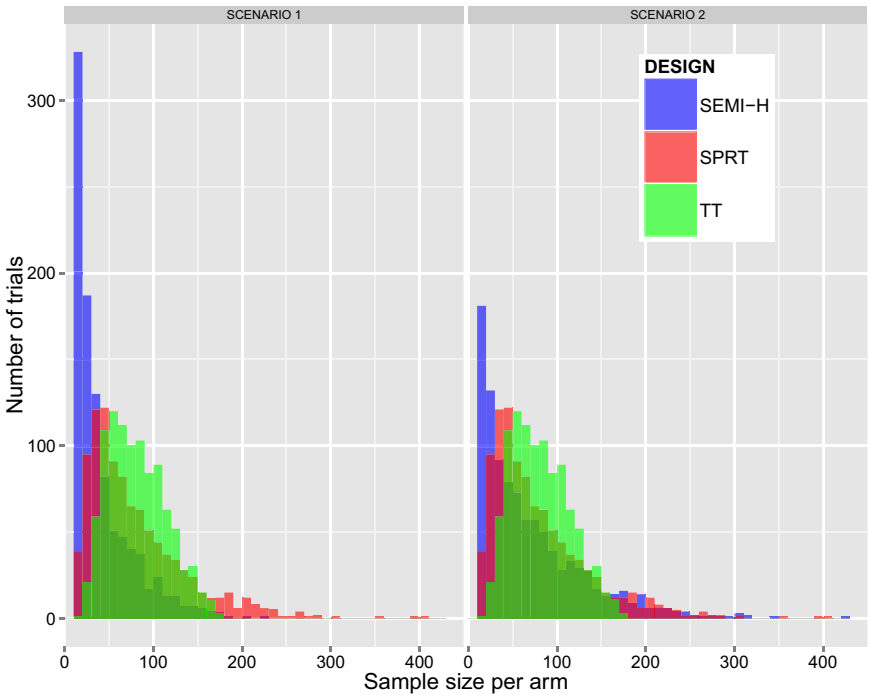

Figure 4 Histograms of the sample sizes per arm obtained at each of the 1,000 clinical trial simulation of the semi-hierarchical (SEMI-H) Bayesian sequential design (blue histogram), sequential probability ratio test (SPRT; red histogram), and the triangular test (TT; green histogram).

effect and associated precisions (**Figure 5**), suggesting that the estimate obtained with a BD (mean μ_{δ_P} of 0.2905) is not significantly influenced by the adult prior distribution under scenario 2.

Similarly to what has been observed when comparing fixed-sample designs, median SS of the SEMI-H (50 patients) is lower than the corresponding value of the SPRT (60 patients) and TT (70 patients). Better performance provided by BDs with respect to this metric can also be deduced from **Figure 4**, where it can be seen that SS histograms of the SPRT ant TT are shifted toward higher sample sizes compared to SEMI-H. However, SEMI-H seems to behave similarly to the SPRT in terms of very late stopping recruitment (i.e., low probabilities exist that the trial goes on very long), as indicated by a 95th percentile in SS distribution of 190.

The right panel of **Figure 5a** reveals that no considerable differences in precision are seen between BSDs and the TT (the frequentist sequential approach with the highest precision), even though Bayesian approaches seem to be slightly more robust to outliers (results not shown). In addition, equivalently to fixed-sample designs, the adult prior distribution does not remarkably influence the estimated effect in pediatrics in the NON-H (mean μ_{δ_P} of 0.3964) and SEMI-H (mean μ_{δ_P} of 0.3939) compared to the SPRT and TT (**Figure 5b**).

In a way, the comparison made under scenario 2 could be considered fairer because type I errors obtained under scenario 2 (around 7–9%) are closer to those obtained in frequentist designs (around 5–7%). On the other hand, the increased type I error rate of BDs observed under scenario 1 is inherently due to the inclusion of a positive adult study and should be accepted as such.

With respect to the extrapolation of adult results to pediatric trials, Hlavin *et al.*[18] proposed a statistical framework to quantitatively accommodate the uncertainty about the assumptions on the similarity between the adult and pediatric population by enlarging the significance level of the pediatric trial based on experts skepticism. Although in the framework by Hlavin *et al.*,[18] Bayesian arguments are applied to calibrate the increase in the significance level, their approach is frequentist by nature and it is does not provide a clear way on how to quantitatively derive the skepticism factor on the basis of the expected similarities/differences between the two populations. Nevertheless, the condition of no skepticism can be translated into the condition $v = 0$ of Schoenfeld *et al.*[5] (children and adults respond in the same way to the drug under study and no pediatric efficacy trial would be needed). Similarly, full skepticism can be converted into $v -> \infty$. Accordingly, for values of v approaching the standard frequentist method ($v \sim 2$), the type I error we obtained corresponds to the adjusted α value proposed by Hlavin *et al.*[18] in case of full skepticism, that is, 0.05 (i.e., no adjustment).

Although we did not explicitly consider advantages and disadvantages of the investigated designs concerning the evaluation of dose regimens in the pediatric population, some general properties can still be outlined. Because BSD and BD are parallel in nature, they differ solely in terms of SS with respect to the estimation of the PK and/or PK-PD in children; as a result, on average, BD is expected to provide more precise PK/PK-PD estimates compared to BSD.

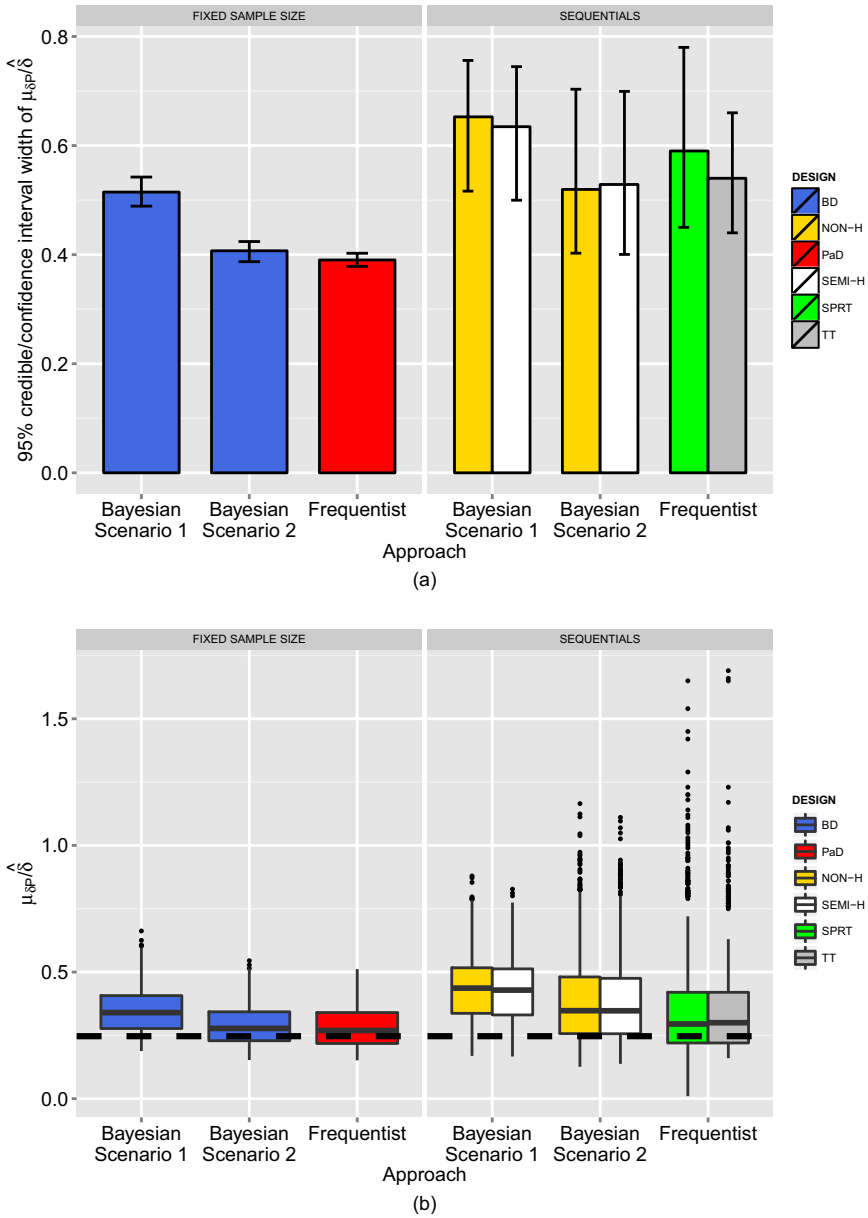

Figure 5 Bar chart plot of the 95% credible (Bayesian) and confidence (frequentist) interval width of treatment difference estimates in pediatrics (μ_{δ_P} for Bayesian designs (BDs) and δ for frequentist ones) (**a**) and boxplot of μ_{δ_P} and $\hat{\delta}$ (**b**) obtained at each of the 1,000 clinical trial simulation of the BD (blue bar), parallel design (PaD; red bar), Bayesian sequential design with a non-hierarchical (NON-H; yellow bar), and a semi-hierarchical (SEMI-H; white bar) framework, sequential probability ratio test (SPRT; green bar), and triangular test (TT; gray bar). The upper and lower "hinges" in subfigure (**a**) correspond to the first and third quartiles of 95% credible/confidence intervals widths. The dashed horizontal black line represents the pediatric treatment effect difference between topiramate and placebo obtained from the pharmacokinetic-pharmacodynamic (PK-PD) model (0.2467).

Similarly, when significant weight is given to prior information on adult treatment effect, BDs would lead to estimates with poorer precision in contrast with frequentist ones; however, if also the PK/PK-PD in children is expected to be similar to that observed in adults, prior information on adult PK/PK-PD parameters can be leveraged to improve the precision of the estimates in the pediatric population and to ultimately provide an optimal dose selection.[6]

Regulatory endorsements on the use of BDs for pediatric trials are present: The European Medicines Agency suggests using Bayesian approaches in pediatric investigation plans,[3] whereas the US Food and Drug Administration published a "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials."[19] Nevertheless, few examples can be found in the literature. According to Gönen,[20] the barriers to entry are many, but three stand out: prior, software, and motivation, where motivation seems to be the major one. We would also add tradition as an additional hurdle, which is anyway related to motivation. Pediatric trials call for innovation and may therefore offer the opportunity to overcome

these motivational issues and increase usage of Bayesian approaches.

It has to be pointed out that our analysis is based on the effect of TPM in children with partial onset seizures refractory to their current antiepileptic treatment, and the extrapolation of our results to different compounds/diseases/subpopulations should be further explored.

In conclusion, in this work, we provided a pharmacometric framework able to formalize PK-PD based CTS for BDs in pediatric trials using prior information from adult data, thereby allowing the investigation of the influence of a specific study design on success/failure of a pediatric trial.

With respect to the selection of a particular design, if prior information is available from adult studies but children are expected to respond substantially different from adults (scenario 2), the performance of frequentist and Bayesian approaches can be assumed comparable, with slight advantages for the latter.

However, when the pediatric population is expected to respond similarly to adults (scenario 1), BDs would allow smaller, shorter, more reliable, and more efficient trials in children. Among BDs, those of sequential nature, irrespective of their level of hierarchy, seem to require lower SS compared to the BD if larger treatment effects are expected, and could therefore represent an appealing option for trials in very small populations.

Acknowledgments. The research leading to these results has received funding from the European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreement 261060 (Global Research in Paediatrics–GRiP network of excellence, www.grip-network.org).

Conflict of Interest. G.S., P.B., A.C., and P.M. have no conflict of interest. M.C. was a full-time employee of Pfizer Ltd at the time of submission.

Author Contributions. M.C., G.S., A.C., and P.M. wrote the manuscript. M.C., G.S., P.B., A.C., and P.M. designed the research. M.C., G.S., and P.M. performed the research. M.C., G.S., and P.M. analyzed the data.

1. Smania, G., Baiardi, P., Magni, P. & Cella, M. Model-based assessment of alternative study designs in pediatric trials. Part I: frequentist approaches. *CPT Pharmacometrics Syst. Pharmacol.*
2. Berry, D.A. Bayesian clinical trials. *Nat. Rev. Drug. Discov.* **5**, 27–36 (2006).
3. European Medicines Agency (EMA). Concept paper on extrapolation of efficacy and safety in medicine development. Human Medicines Development and Evaluation. <http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/04/WC500142358.pdf> (2012).
4. Goodman, S.N. & Sladky, J.T. A Bayesian approach to randomized controlled trials in children utilizing information from adults: the case of Guillain-Barre syndrome. *Clin. Trials.* **2**, 305–310; discussion 364-378 (2005).
5. Schoenfeld, D.A., Hui, Z. & Finkelstein, D.M. Bayesian design using adult data to augment pediatric trials. *Clin. Trials.* **6**, 297–304 (2009).
6. Cella, M., Gorter de Vries, F., Burger, D., Danhof, M. & Della Pasqua, O. A model-based approach to dose selection in early pediatric development. *Clin. Pharmacol. Ther.* **87**, 294–302 (2010).
7. Berry, D.A. Interim analyses in clinical trials: classical vs. Bayesian approaches. *Stat. Med.***4**, 521–526 (1985).
8. Cronin, K.A., Freedman, L.S., Lieberman, R., Weiss, H.L., Beenken, S.W. & Kelloff, G.J. Bayesian monitoring of phase II trials in cancer chemoprevention. *J. Clin. Epidemiol.* **52**, 705–711 (1999).
9. Thall, P.F. & Simon, R. Practical Bayesian guidelines for phase IIB clinical trials. *Biometrics* **50**, 337–349 (1994).
10. Chen, C. & Chaloner, K. A Bayesian stopping rule for a single arm study: with a case study of stem cell transplantation. *Stat. Med.* **25**, 2956–2966 (2006).
11. Kashiwabara, K., Matsuyama, Y. & Ohashi, Y. A Bayesian stopping rule for sequential monitoring of serious adverse events. *Ther. Innov. Regul. Sci.* **48**, 444–452 (2014).
12. Guédé, D. *et al.* Bayesian adaptive designs in single ascending dose trials in healthy volunteers. *Br. J. Clin. Pharmacol.* **78**, 393–400 (2014).
13. Girgis, I.G. *et al.* Pharmacokinetic-pharmacodynamic assessment of topiramate dosing regimens for children with epilepsy 2 to <10 years of age. *Epilepsia* **51**, 1954–1962 (2010).
14. U.S. Food and Drug Administration. Topomax (topiramate): label change – risk for development of cleft lip and/or cleft palate in newborns. <http://www.fda.gov/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm245777.htm> (2012).
15. European Medicines Agency (EMA). Topomax package leaflet. <http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/referrals/Topamax/human_referral_000170.jsp> (2009).
16. Gsponer, T., Gerber, F., Bornkamp, B., Ohlssen, D., Vandemeulebroecke, M. & Schmidli, H.A. Practical guide to Bayesian group sequential designs. *Pharm. Stat.* **13**, 71–80 (2014).
17. Lee, J.J. & Chu, C.T. Bayesian clinical trials in action. *Stat. Med.* **31**, 2955–2972 (2012).
18. Hlavin, G., Koenig, F., Male, C., Posch, M. & Bauer, P. Evidence, eminence and extrapolation. *Stat. Med.* **35**, 2117–2132 (2016).
19. U.S. Food and Drug Administration. Guidance for the use of Bayesian statistics in medical device clinical trials. <http://www.fda.gov/RegulatoryInformation/Guidances/ucm071072.htm> (2010).
20. Gönen, M. Bayesian clinical trials: no more excuses. *Clin. Trials.* **6**, 203–204 (2009).

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (http://www.wileyonlinelibrary.com/psp4)