## ORIGINAL ARTICLE

# Model-Based Assessment of Alternative Study Designs in Pediatric Trials. Part I: Frequentist Approaches

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

Alternative designs can increase the feasibility of pediatric trials when compared to classical parallel designs (PaD). In this work we present a model-based approach based on clinical trial simulations for the comparison of PaD with the alternative sequential, crossover, and randomized withdrawal (RWD) designs. Study designs were evaluated in terms of: type I and II errors, sample size per arm (SS), trial duration (TD), treatment exposures, and parameter estimate precision (EP). The crossover requires the lowest SS and TD, although it implies higher placebo and no treatment exposures. RWD maximizes exposure to active treatment while minimizing that to placebo, but requires the largest SS. SS of sequential designs can sometimes be smaller than the crossover one, although with poorer EP. This pharmacometric framework allows a multiscale comparison of alternative study designs that can be used for design selection in future pediatric trials.
*CPT Pharmacometrics Syst. Pharmacol.* (2016) **5**, 305–312; doi:10.1002/psp4.12083; published online 1 June 2016.

## Study Highlights

**WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?**
☑ Randomized controlled trials in pediatrics faces several difficulties and standard parallel designs may not always be feasible in this population. Crossover, randomized withdrawal, and sequential designs offer advantages that could potentially help to streamline pediatric drug development.
**WHAT QUESTION DID THIS STUDY ADDRESS?**
☑ This study presents a PK-PD based clinical trial simulation framework with the aim of assessing the performance of alternative study designs in pediatric trials across different metrics of comparison.

**WHAT THIS STUDY ADDS TO OUR KNOWLEDGE**
☑ Based on the peculiarities and the clinical context of the trial, alternative designs may be preferable to the standard parallel design. This study provides instruments for informed decision-making in pediatric trials.
**HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY AND THERAPEUTICS**
☑ A methodological approach based on PK-PD-based clinical trial simulation will allow optimizing the design of future pediatric trials, ultimately increasing their feasibility.

Pediatric drug development faces several difficulties due to ethical, practical, and financial considerations. Despite the EU Pediatric Regulation (EC No. 1901/2006),[1] the US Best Pharmaceuticals for Children Act,[2] and the US Pediatric Research Equity Act[3] partially saved children from their role of therapeutic orphans by facilitating the execution of pediatric clinical trials, a number of obstacles still remain in providing children with safe and effective drugs.[4–6]

Consequently, the design and analysis of pediatric clinical trials necessitate the most efficient and informative analytical methods.[7] The gold standard method for assessing the efficacy and safety of a new drug in patients is the randomized controlled trial (RCT), which minimizes bias and provides a clear and reliable understanding of the risk/benefit ratio of a new experimental treatment. Conventional confirmatory RCTs are mostly executed using a parallel-arms design with hundreds or even thousands of patients enrolled. Since the number of patients that can be enrolled in pediatric studies is limited, trials of such sizes are often unfeasible. Scientifically,

clinically, and logistically plausible alternatives to the classical parallel design are therefore needed.[8]

The aim of this work is to compare the performance of the classical parallel design (PaD) with that of the alternative crossover, randomized withdrawal and sequential designs by means of pharmacokinetic–pharmacodynamic (PK-PD)-based clinical trial simulation (CTS). Bayesian approaches are investigated in the Part II article.[9] Known advantages and disadvantages of the study designs that will be evaluated in the present work when compared to the standard parallel design are outlined in **Table 1**.

A published pediatric PK-PD model of topiramate (TPM)[15] was used as a paradigm for CTS in epileptic children. Designs were evaluated in terms of: type I and type II errors; sample size per arm; total trial duration; relative extent of placebo, active treatment and no-treatment exposure (due to periods during which the patients do not take neither TPM nor placebo, e.g., baseline and/or washout periods), and precision of treatment difference estimate. For some of the

[1]Consorzio per Valutazioni Biologiche e Farmacologiche, Pavia, Italy; [2]Dipartimento di Ingegneria Industriale e dell'Informazione, Università degli Studi di Pavia, Pavia, Italy; [3]Direzione Scientifica Centrale, Fondazione Salvatore Maugeri, IRCCS, Pavia, Italy; [4]Department of Clinical Pharmacology, Global Clinical Development, Chiesi Farmaceutici, Parma, Italy. *Correspondence: M Cella (mcella@cvbf.net)

**Table 1** Pros and cons known from the literature of alternative study designs considered in the present work when compared to the standard parallel design

| Study design | Pros | Cons | References |
|---|---|---|---|
| Crossover | • Smaller sample size<br>• All participants receive treatment | • Carryover effect<br>• Longer duration<br>• Not suitable if the disease is not stable over time<br>• Not suitable in case of treatment with permanent effect<br>• All participants receive placebo<br>• Need for washout period | [10] |
| Randomized Withdrawal | • Subjects continue receiving study drug only if they respond to it<br>• Lower exposure to placebo<br>• Enrichment of study population | • Carryover effect<br>• Treatment effect estimate is biased towards responders<br>• Suitable only for stable chronic diseases<br>• Ethical concerns with depriving patients of the benefit they had already obtained from the active drug | [8,11,12] |
| Sequential designs | • Allows to terminate a trial when evidence has emerged that one treatment is clearly either superior or inferior to the other<br>• Sample size is on average smaller | • Treatment outcomes should be available quickly in relation to patients recruitment rate<br>• Maximum sample size can be larger<br>• Increased logistic complexity | [13,14] |

investigated designs, part of these measures have been computed analytically without the need for CTS.

CTS has been successfully used in pediatrics to help trial design, not only for dose selection,[16–20] but also to set other trial features such as number of dose groups and number of patients per group.[21–24] However, to the best of our knowledge, no attempts were made on the simultaneous investigation of a battery of alternative designs different from parallel/ crossover. Moreover, comparisons are normally built upon purely statistical criteria such as sample size and mean square error of estimates, whereas we hereby propose the additional use of total trial duration and treatments exposure to evaluate the overall performance of a particular design.

## METHODS
### Case study: PK-PD model of topiramate adjunctive therapy in children with epilepsy
We first reviewed the literature in order to identify a pediatric PK-PD model suitable for our analysis. Among the few models that resulted from the literature search, many were inadequate because of insufficient details to allow the use of the model in a CTS setting, lack of the PK component, unsatisfactory model evaluation, and modeling of a safety PD measure rather than an efficacy one. The final choice was a PK-PD model of TPM in pediatric patients between 2 and 10 years of age with partial-onset or primary generalized tonic–clonic seizures.[15] The model is described in **Supplementary Material 1.**

### Study designs description
All study designs presented in this work are alternative implementations of a two-arm RCT. Patients in the control group received placebo (i.e., their current antiepileptic treatment plus placebo), while patients in the treatment group received 7 mg/kg b.i.d. of TPM (i.e., their current antiepileptic treatment plus TPM), that is the average US Food and Drug Administration/European Medicines Agency (FDA/EMA)-recommended TPM dosage regimen for the adjunctive treatment of epileptic children.[25,26] The clinical endpoint of the trial was the log-transformed translated percent reduction in seizure frequency from baseline (i.e., Y, see **Supplementary Material 1**). Coherently with the model and in agreement with previous findings, we set the length of the baseline and the treatment phase at 1 month each, for an overall duration of the trial per child ($\tau$) of 2 months.[27]

In order to design the studies, an initial estimate of both the improvement of TPM over placebo ($\delta = \mu_{PCB}\text{-}\mu_{TPM}$, where $\mu_{PCB}$ and $\mu_{TPM}$ are the expected placebo and TPM responses in terms of Y) and the variability of Y ($\sigma$) have to be formulated. $\sigma$ was derived from the original publication and set to 0.7517, whereas Monte Carlo methods were used to compute $\delta$ from $10^6$ samples, leading to a value of $\delta = 0.2467$. An improvement of 0.2467 in the Y scale corresponds to approximately a 19% further decrease in seizure reduction for TPM 7 mg/kg against placebo, considering an average placebo seizure reduction of 21.5% (obtained from the PK-PD model). The superiority of TPM over placebo was assessed through standard one-sided statistical testing on the null hypothesis of no treatment difference H$_0$: $\delta < 0$ with 5% significance (i.e., $\alpha = 0.05$) and 80% power (i.e., $\beta = 0.20$).

*Parallel design (PaD).* In a two-arm PaD, patients are randomized into two parallel groups to receive either placebo or TPM, with the number of patients to be randomized in each group fixed *a priori*. In agreement with the PK-PD model, we assumed responses to be normally distributed with the same variance $\sigma^2$ in the TPM and placebo arm. Accordingly, we used normal-approximation for sample size calculation for a one-sided Student's *t*-test to obtain the number of patients to be enrolled in each group $n = 2\left[\frac{(z_{1-\alpha}+z_{1-\beta})\cdot\sigma}{\delta}\right]^2$, with $z_x$ being the x-th quantile of the standard normal distribution.

*Crossover design (XD).* In an XD, patients are randomized into two treatment sequences, one where they receive first TPM and then placebo, and one where they receive first placebo and then TPM. The length of the washout period between the two treatment sequences was set to 1 month, in accordance with previous crossover studies in pediatric epileptic patients.[28]

Sample size calculation for the XD was adapted from Wellek and Blettner.[10] Similarly to the PaD, we assumed responses to be normally distributed with the same variance $\sigma^2$ in patients receiving TPM and placebo. In particular, in order to obtain the number of patients to be randomized in each sequence, the following formula (assuming normal-approximation of t distribution) was used: $n = 2\left[\frac{(z_{1-\alpha} + z_{1-\beta}) \cdot \sigma}{2\delta}\right]^2 [2(1-\rho)]$, where $\rho$ is the correlation of Y between the two periods of the XD ($\rho$ can be thought of as the proportion of PD Between Subject Variability (BSV) contained in $\varepsilon$: if $\rho = 0$ then all the variability contained in $\varepsilon$ is intraindividual variability (IIV), if $\rho = 1$ then all the variability contained in $\varepsilon$ is BSV). In order to evaluate the sensitivity of design performance with respect to such parameter, simulations were carried out for $\rho = 0$, 0.25, 0.5, 0.75.

*Randomized withdrawal design (RWD).* In an RWD, after an initial open-label period in which all patients receive TPM, only patients who positively responded to TPM (defined as patients whose percentage seizure reduction from baseline is greater than the corresponding average placebo response) enter the double-blind phase and are randomized to receive either placebo or TPM, whereas the nonresponders discontinue the trial. The same washout period and correlation $\rho$ defined in the XD were assumed in the RWD between the open-label and double-blind phase.

In order to maintain the desired statistical properties of the analysis, the sample size of the RWD in the double-blind phase (whose collected measures will be subjected to statistical testing) should be similar to the PaD one. Therefore, an initial estimate of the percentage of responders ($\theta$) is needed in order to obtain the total sample size at the open-label phase, which is defined as $4\left[\frac{(z_{1-\alpha} + z_{1-\beta}) \cdot \sigma}{\delta}\right]^2 \cdot \frac{1}{\theta}$. Following the same procedure used to obtain $\delta$, PK-PD simulations allowed us to derive an estimate of the responder rate $\theta$ of 0.627, suggesting that about 62.7% of children have a response to TPM greater than the average placebo response.

*Group sequential designs: sequential probability ratio test (SPRT) and triangular test (TT).* In group sequential designs, 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. Several statistical approaches have been proposed for the design and analysis of group sequential designs (e.g., O'Brien–Fleming method[29] and Pocock method[30]). In our work, we considered two alternative implementations of group sequential designs, namely, the SPRT and TT. Although such designs have been rarely applied, they appear to have favorable properties for pediatric trials.[31] The statistical framework for these two
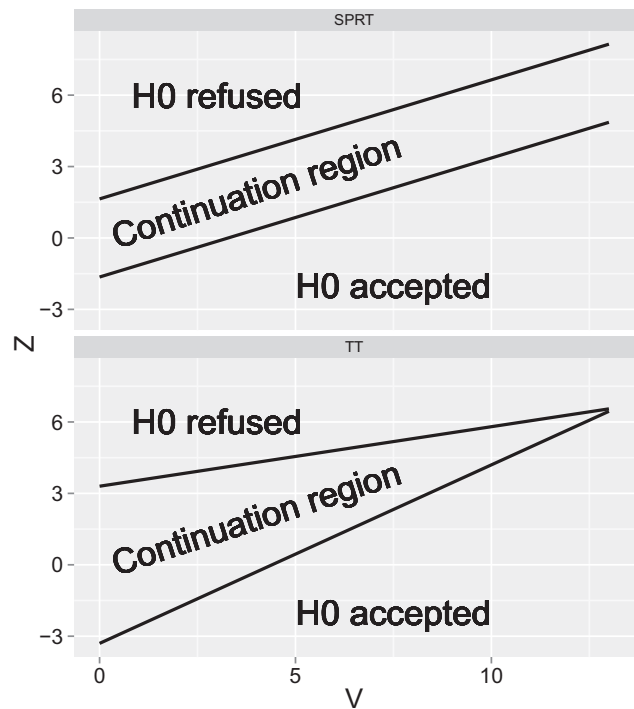


**Figure 1** Example of acceptance/rejection boundaries of the sequential probability ratio test (SPRT, upper panel) and the triangular test (TT, lower panel) for $\delta = \sigma = 1$ and $\alpha = \beta = 0.05$. During the trial the value of the Z statistics computed at each analysis is plotted against the associated V-value, building up a path on the Z-V plane. If such a path crosses the upper boundary, $H_0$ is refused; if it crosses the lower boundary, $H_0$ is accepted; if it stays within the continuation region, patient recruitment goes on.

designs was adapted from Whitehead[32] (a formal presentation can be found in **Supplementary Material 1**). The SPRT and TT are also known as boundary methods since, at each interim analysis, a sample statistics Z (which can be thought of as the accumulated evidence of $\delta$) is plotted against a second sample statistics V (which can be thought of as the amount of information about $\delta$ contained in Z), and when the value of Z exits a so-called continuation region delimited by two boundaries in the V-Z plane, $H_0$ is either accepted or refused (**Figure 1**). The two methods differ for the equations of the boundaries: in the SPRT these are parallel and the continuation region is open, while in the TT they converge, defining a close continuation region. On the one hand, the TT may thereby appear more relevant *a priori*, because the sample size could theoretically be infinite by using the SPRT. On the other hand, sample size reductions in the case of clear evidence of efficacy/futility are larger with the SPRT when compared to the TT.

**Study designs simulation**
The simulation of each design was based on the following stepwise procedure:

**Table 2** Metrics $\hat{\alpha}$: type I error; $\hat{\beta}$: type II error; SS: sample size per arm; $SS_{50}$: median sample size per arm; $SS_{75}$, $SS_{90}$, $SS_{95}$: 75th, 90th and 95th percentiles of sample size distribution, respectively; TD: total trial duration obtained from clinical trial simulation for each of the simulated designs and for different values of $\rho$. PaD: parallel design; XD: crossover design; RWD: randomized withdrawal design; SPRT: sequential probability ratio test; TT: triangular test. ER: enrollment rate.*SS reported in the table is the one of the double-blind phase, SS as defined in **Supplementary Material 5** (i.e., as half of the sample size in the open-label phase) is 184.

| Design Metric | PaD | XD | | | | RWD* | | | | SPRT | TT |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\rho = 0$ | $\rho = 0.25$ | $\rho = 0.5$ | $\rho = 0.75$ | $\rho = 0$ | $\rho = 0.25$ | $\rho = 0.5$ | $\rho = 0.75$ | | |
| $\hat{\alpha}$ (%) (95% CI) | 5.1 (3.7–6.5) | 4.7 (3.4–6.0) | 5.4 (4.0–6.8) | 5.2 (3.8–6.6) | 4.4 (3.1–5.7) | 5.4 (4.0–6.8) | 4.9 (3.6–6.2) | 4.9 (3.6–6.2) | 4.2 (3.0–5.4) | 7.0 (5.4–8.6) | 5.9 (4.4–7.4) |
| $\hat{\beta}$ (%) (95% CI) | 21.2 (18.7–23.7) | 19.1 (16.7–21.5) | 22.8 (20.2–25.4) | 22.5 (19.9–25.0) | 25.0 (22.3–27.7) | 13.4 (11.3–15.5) | 13.3 (11.2–15.4) | 10.7 (8.8–12.6) | 7.5 (5.9–9.1) | 18.3 (15.9–20.7) | 20.3 (17.8–22.8) |
| E[SS] (patients) | 115 | 58 | 44 | 29 | 15 | 116 | 116 | 116 | 116 | 75 | 77 |
| $SS_{50}$ (patients) | 115 | 58 | 44 | 29 | 15 | 115 | 116 | 115 | 116 | 60 | 70 |
| $SS_{75}$ (patients) | 115 | 58 | 44 | 29 | 15 | 118 | 118 | 118 | 119 | 100 | 100 |
| $SS_{90}$ (patients) | 115 | 58 | 44 | 29 | 15 | 121 | 121 | 121 | 121 | 150 | 120 |
| $SS_{95}$ (patients) | 115 | 58 | 44 | 29 | 15 | 122 | 123 | 123 | 123 | 181 | 140 |
| MEDIAN TD (months) ER = 4 patients/month | 59.5 | 33.0 | 26.0 | 18.5 | 11.5 | 96.0 | 96.0 | 96.0 | 96.0 | 32 | 40 |
| ER = 10 patients/month | 25.0 | 15.6 | 12.8 | 9.8 | 7.0 | 48.0 | 48.0 | 48.0 | 40.8 | 14 | 16 |

1. Patients population was simulated in terms of weight and age (as these were the only two significant covariates of the PK model) using weight-for-age tables from World Health Organization statistics[33] (a detailed description can be found in **Supplementary Material 1**).
2. For each virtual child, the individual $C_{min}$ was extracted from 500 PK profiles generated by using NONMEM v. 7.2.0[34] (see **Supplementary Material 2** for NONMEM code used).
3. Patients to be enrolled following the design-specific rules were randomly selected from the population simulated at Step 1 and, for each child, a single $C_{min}$ was further sampled from its 500 values simulated at Step 2. These values were fed into the PK-PD model in order to simulate the clinical endpoint. For subjects randomized to the control arm, $C_{min}$ was set to zero.
4. The study design-specific statistical analysis was applied to the simulated endpoints.

Steps 3 and 4 were reiterated 1,000 times per study design, i.e., 1,000 CTS were performed. A visual description of this stepwise procedure can be found in **Figure S1**. R software v. 3.0.1[35] was used for Steps 1 (**Supplementary Material 3**), 3, and 4 (**Supplementary Material 4**).

The following metrics were adopted for designs comparison:

a. *Type I error ($\hat{\alpha}$)*: the proportion of statistical analyses leading to the rejection of $H_0$ when simulating under $H_0$ (i.e., when in the simulations described at Step 3 $\beta_1 = 0$).
b. *Type II error ($\hat{\beta}$)*: the proportion of statistical analyses leading to the acceptance of $H_0$ when simulating under $H_1$ (i.e., when in the simulations described at Step 3 $\beta_1$ is set to its estimated value).
c. *Sample size per arm (SS)*: number of children enrolled in each arm of the trial.
d. *Total trial duration (TD)*: the duration in months of the entire trial as a function of enrollment rate (ER, we considered ER up to 10 patients/month). Although certain features of a trial design (e.g., treatment blinding, inclusion/exclusion criteria) can influence ER,[36] these were not considered in our analysis and we therefore assumed ER to be design-independent.
e. *Average extent of placebo, TPM, and no-treatment exposure*: percentage of exposure to placebo, TPM, and no-treatment relative to total trial exposure.
f. *Treatment difference estimate precision (EP)*: the precision of the estimate of $\delta$ ($\hat{\delta}$) expressed in terms of the width of its 95% confidence interval (CI).

**Supplementary Material 5** contains a comprehensive description of the calculation of metrics above for each of the investigated designs.

## RESULTS
### Type I and type II errors
$\hat{\alpha}$ and $\hat{\beta}$ are close to their predetermined levels of 5% and 20% for all designs except for the RWD with $\rho > 0$, where $\hat{\beta}$ appears to decrease when correlation increases (**Table 2**).This increasing power with increasing $\rho$ is due to a decrease in the variability of responses in the double-blind phase of the study, further given by the fact that drug effect is evaluated in a specific subset of the pediatric population (i.e., those who respond to TPM).
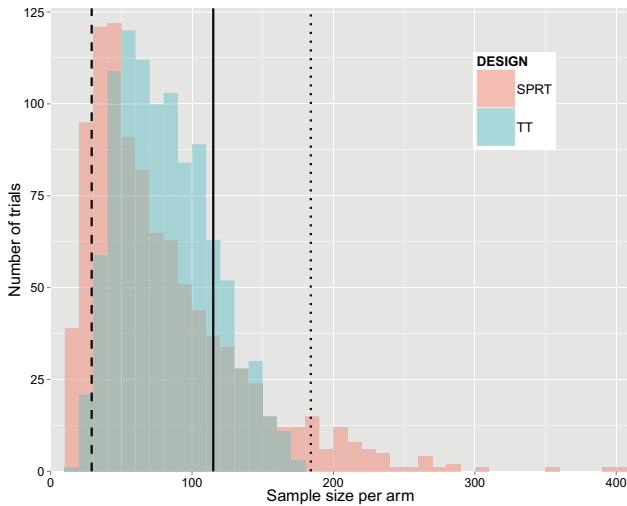
**Figure 2** Histograms of the sample sizes per arm obtained at each of the 1,000 clinical trial simulations of the sequential probability ratio test (SPRT, pink histogram) and the triangular test (TT, blue histogram). Black vertical lines indicate the sample sizes per arm of the parallel design (solid line), crossover design with $\rho = 0.5$ (dashed line), and randomized withdrawal design (dotted line).

Sequential designs show a slightly higher $\hat{\alpha}$, although for the TT the target value of 5% is contained in the 95% CI of the estimated $\hat{\alpha}$. Conversely, $\hat{\alpha}$ of the SPRT is significantly higher than 5%.

### Sample size per arm
**Table 2** shows that the XD with $\rho = 0.75$ leads to an SS of 15 children, which is the minimum SS among all designs. Even for lower values of $\rho$ the XD requires a lower SS
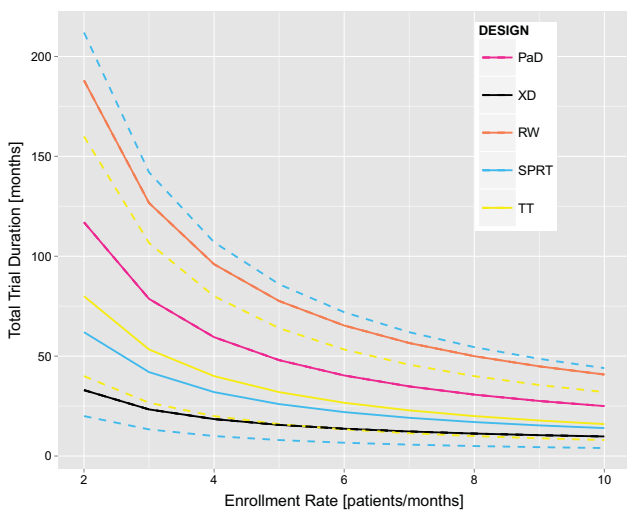


**Figure 3** Total trial duration as a function of enrollment rate for each of the investigated designs (PaD: parallel design; XD: crossover design with $\rho = 0.5$; RWD: randomized withdrawal design with $\rho = 0.5$; SPRT: sequential probability ratio test; TT: triangular test). Solid lines represent the median duration whereas dashed lines depict 95% prediction intervals.
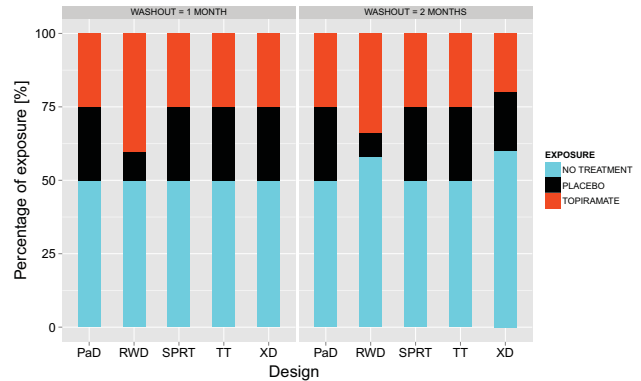


**Figure 4** Percentage of exposure to no-treatment (cyan bar), placebo (black bar), and topiramate (red bar) relative to total trial exposure for each of the investigated designs (PaD: parallel design; XD: crossover design; RWD: randomized withdrawal design; SPRT: sequential probability ratio test; TT: triangular test) for a washout of 1 (left panel) and 2 (right panel) months. For the XD and RWD this metric is equal across all values of $\rho$.

compared with other designs, and when the PD BSV is negligible ($\rho = 0$) SS of XD is half of the SS in the PaD.

SS for sequential designs is not deterministic. The histograms of the SS obtained at each simulation of the two sequential designs are depicted in **Figure 2**. It can be seen that on average both the SPRT and TT requires fewer patients than the PaD (around 76 per arm).

Finally, the estimated probability of terminating the sequential designs with an SS greater than the PaD is fairly low (19% for the SPRT and 13.9% for the TT).

### Total trial duration
TD reflects the required SS: the higher the sample size, the higher the duration (for a given ER). Accordingly the XD has the lowest median TD among all investigated designs (**Figure 3**). The same may not be true for sequential designs however, wherein the time needed before obtaining the primary endpoint for the sequential analysis can remarkably increase TD. Interestingly, the results show that for both pessimistic (4 patients/months) and optimistic (10 patients/months) enrollment rates, this was not the case (**Table 2**). In fact, the median TDs of SPRT and TT are lower than TD of the PaD, suggesting that for this magnitude of ER, TD reduction due to a lower SS outweighs the TD increasing that could have been arisen because of sequential enrollment.

### Extent of placebo, TPM, and no-treatment exposure
The left panel of **Figure 4**, which quantifies the extent of exposure to placebo (black bar), TPM (red bar), and no-treatment (cyan bar) for each of the investigated designs in terms of the proportion of exposure relative to total trial exposure, clearly shows that the RWD allows minimizing exposure to placebo while maximizing exposure to TPM, no-treatment exposure being equal at 50% for all designs.

As to no-treatment exposures, these are comparable among all designs. However, if a washout of 2 months had been required, the percentage of exposure to no-treatment
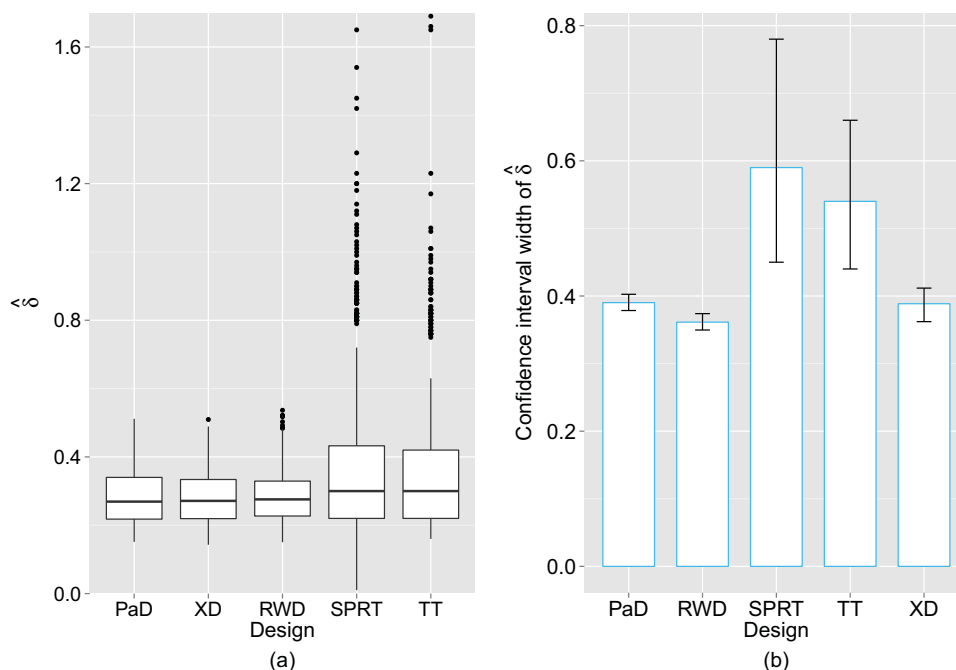
**Figure 5** Boxplot of treatment difference estimates ($\hat{\delta}$) (**a**) and bar chart plot of the median confidence interval width of $\hat{\delta}$ (**b**) obtained at each of the 1,000 clinical trial simulation of the parallel design (PaD), crossover design with $\rho = 0.5$ (XD), randomized withdrawal design with $\rho = 0.5$ (RWD), sequential probability ratio test (SPRT) and triangular test (TT). Nine simulations in the SPRT and one in the TT were not able to compute $\hat{\delta}$ and its precision because of negative values of V due to very early stopping (i.e., final sample size per arm = 10 patients) and were therefore excluded from the figure.

in the XD and RWD would have risen to 60% and 58%, respectively (right panel of **Figure 4**).

**Treatment difference estimate ($\hat{\delta}$) precision**
Although the value of $\hat{\delta}$ was similar for all designs (**Figure 5a**), its precision may substantially vary across them. In fact, **Figure 5b** shows that the sequential designs can lead to precisions much lower than those obtained with a PaD or RWD in terms of width of 95% CI. This is partly explained by the number of samples used to compute the estimate: the SPRT and TT are those designs that allow on average to keep the sample size low, thereby increasing the standard error of $\hat{\delta}$ and consequently the width of its 95% CI.

## DISCUSSION

The use of alternative study designs in pediatric trials can significantly improve their feasibility by reducing their sample size and duration and by increasing their acceptability.

Examples on the use of PK-PD-based CTS for the assessment of trial design performance exist,[21,22] although no attempts have been made on simultaneously exploring alternative study designs such as the RWD and the sequential designs. Investigations of sequential designs by CTS can also be found in the literature[37,38]; however, these are based on purely statistical models, thereby neglecting the PK-PD component. As a result, the influence of patient demographics and dose regimens on trial performance could not have been explicitly taken into account. The pres-

ent work provides a pharmacometric-based framework for a multidimensional comparison of alternative study designs.

Overall, the outcomes of our analysis are in line with the known pros and cons of the investigated designs introduced in **Table 1**. The results clearly show that for a pediatric trial the XD, irrespective of the value of $\rho$, allows minimizing the SS required while maintaining desired type I and type II errors. The minimization of SS translates also into a very low TD. However, the XD may not be easily accepted by parents/children because the washout period implies that children have to spend a greater amount of time without taking any treatment when compared to other designs (**Figure 4**). In addition, despite that all children enrolled in the trial will certainly receive the active treatment, they will certainly receive placebo as well, posing further ethical issues. Finally, although we assumed negligible carryover effects, these have to be considered when designing a pediatric trial with a crossover scheme because they can eventually compromise the analysis and interpretation of the results.[10,39]

The RWD ensures that all children enrolled in the trial will receive the new treatment and those not responding to the treatment will be quickly withdrawn from the trial. As shown in **Figure 4**, for a washout of 1 month, the percentage of exposure to TPM in the RWD is 40%, compared with 25% of other designs. At the same time, the percentage of exposure to inactive treatment (placebo) is less than half that of other designs. Such properties are still valid if we double the washout period. These parent/patient-friendly features of the RWD, along with an acceptable level of scientific rigor, contributed to its increased popularity in the design of juvenile arthritis trials.[11,12]

Nonetheless, our results suggest that the RWD would require a higher SS compared to other designs (**Figure 2**) and, consequently, a higher TD (**Figure 3**). However, if it is reasonable to assume that patients response to treatment does not remarkably change between the open-label and double-blind period (i.e., $\rho = 0.75$), the RWD leads to a greater power (92.5%, **Table 2**). From an SS perspective, by maintaining the type II error to $\sim$20%, the SS of the RWD would drop to values around that of PaD (115, results not shown). Because in RWD a slightly different population is studied compared to the other designs, a different $\hat{\delta}$ is typically expected. **Figure 5a** shows that this was not the case in our analysis. This is primarily due to the fact that a certain patient is classified as a responder in the open-label phase mainly because a large $\varepsilon$ was sampled for that patient, and not because that patient had an increased exposure to TPM in terms of $C_{min}$. This implies that: 1) for large values of $\rho$ ($> \sim 0.5$), patients in the double-blind phase have a high response in both the placebo and TPM arm, which prevents an increase of $\hat{\delta}$ (e.g., for $\rho = 0.75$ mean $\hat{\delta}$ turned out to be around 0.270 in the RWD, compared to the value of 0.277 observed in the PaD); 2) for low values of $\rho$ ($< \sim 0.5$), the effect of a larger $C_{min}$ on $\hat{\delta}$ is partially masked by high (or low) values of $\varepsilon$ (mean $\hat{\delta}$ was estimated around 0.286 in the RWD, compared to the value of 0.277 observed in the PaD). Consequently, our results are likely to be observed in a RWD in patients with a large placebo response (1) or when the magnitude of the BSV is negligible compared to IIV (2).

Sequential designs are of great interest for pediatric trials, essentially because they allow an early stop for efficacy or futility. Previous pediatric sequential trials have shown a median SS decrease of 35% compared to standard PaDs.[31] Our results confirm that on average both the SPRT and the TT determine an SS reduction compared to the PaD (between 33% and 50%), without compromising the desired statistical properties (although $\hat{\alpha}$ seems to be slightly higher, in agreement with earlier findings[37]).

Moreover, our simulations show that the SPRT and TT have a 13.4% and 2.2% probability of demonstrating drug efficacy with an SS lower than 21, respectively. Accordingly, TD for an enrollment rate of just two patients/month would fall to 20 months. Despite that the treatment effect estimate precision associated with these low SS cannot be considered acceptable, these designs may be of interest when very limited subjects can be recruited. On the other hand, since in sequential designs the SS is not fixed *a priori*, the final SS and TD of the SPRT and TT may turn out to be greater than the one that would have been required by a fixed sample size approach. This is demonstrated by a 90th percentile of SS distribution greater than the PaD SS (**Table 2**). Within sequential designs, the TT appears to outperform the SPRT in the unfortunate scenario of a late study termination, with a 90th percentile in SS distribution of 30 patients lower than the SPRT. These results are in line with those from Sebille and Bellissant.[37]

The added value of our analysis compared to that of Sebille and Bellissant is that by using PK-PD-based CTS we are able to contextualize the analysis in the clinical condition under study and to investigate the impact of PK variability and patient characteristics on the possible results of the trial. CTS becomes then a tool for a sound evaluation of candidate designs by enabling, for example, to assess the impact of patient population characteristics on the probability of terminating a sequential trial with an SS greater than that of a traditional design.

We acknowledge that the present framework is underpinned by a robust pediatric PK-PD model that may not always be available at the time of the design of the pediatric trial. If this is the case, extrapolations using an adult PK-PD model in lieu of the pediatric one can be considered, provided that the following conditions can reasonably be assumed[40]: 1) the pathophysiology of the disease is the same between children and adults; 2) either the PK in pediatric patients is known or extrapolation of the PK from adult data is suitable (i.e., differences in PK are explained solely by differences in body weight, a reasonable assumption in children older than 2 years[41]); 3) the PK-PD relationship is similar between the two populations; 4) the response to treatment is assessed in terms of the same PD measure in pediatrics and adults. In general, the degree of uncertainty of assumptions (1–4) should guide the design of the pediatric study. Halvin *et al.*[42] proposed a statistical framework to quantitatively accommodate assumptions uncertainty by enlarging the significance level of the pediatric trial based on experts' skepticism about the expected similarities and differences between the adult and pediatric population. The use of an adult PK-PD model in our framework would implicitly convey a certain magnitude of skepticism (or rather belief) in the extrapolation process. The dependency of trial results and design performance on this magnitude of skepticism can potentially be investigated through the integration of our framework with that proposed by Halvin *et al*. This would ultimately enable researchers to select the study design that best suits their current extrapolation concept.

Since we focused on pediatric efficacy trials, the strengths and weaknesses of study designs with respect to ancillary trial objectives have not been investigated in our analysis. One important aspect relates to the support of dose regimen in the pediatric population. From this point of view, the RWD is expected to be one of the more robust in justifying pediatric dosage because it allows emphasizing the effect of the tested dose compared to placebo by ruling out the confounding element that would be introduced by the randomization of non-responders. In addition, the XD has the favorable property of estimating the true drug effect in each patient, leading to more precise estimates of BSV of a drug effect parameter; nevertheless, its small sample size might jeopardize the reliability of the PK analysis. The same would also apply for sequential designs terminating with a very low sample size. PaD and sequential designs do not exhibit any particular advantage when it comes to supporting dose regimens in the pediatric population, and the only difference between the two is attributable to differences in sample sizes.

Noteworthy, we did not include a dropout model and did not test whether dropout patterns are influenced by the study design; accordingly, any knowledge of differences could lead to a more accurate case-specific evaluation. Moreover, 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.

To conclude, there is no best study design for children with refractory epilepsy that performs better in all the metrics we have monitored. Sequential designs are probably more appealing because they appear to considerably reduce the SS when large effect sizes are expected. This is particularly important if patient recruitment is the primary obstacle, as TD is not inflated by the sequential procedure of the design and low precisions in $\hat{\delta}$ may be tolerated. On the other hand, if major concerns involve the ethical acceptability of the trial, an RWD may be preferable because of the shortened placebo exposure and simultaneously increased exposure to the active treatment, especially if it is reasonable to assume that the individual response to treatment does not significantly change between the open-label and double-blind phase.

In general, pediatric design selection would largely benefit from a pharmacometric approach as the one described in our work, which leverages prior information available and allows to test different "what if" scenarios by assessing the characteristics of the design across multiple levels.

**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 first submission.

**Author Contributions.** G.S., P.M., and M.C wrote the article; G.S., P.B., A.C., P.M., and M.C. designed the research; G.S., P.M., and M.C performed the research.

1. EU. Regulation on medicinal products for paediatric use. Official Journal of the European Union (2006).
2. FDA. Best Pharmaceuticals for Children Act (2002).
3. FDA. Pediatric Research Equity Act (2003).
4. Tishler, C.L. & Reiss, N.S. Pediatric drug-trial recruitment: enticement without coercion. *Pediatrics* **127**, 949–954 (2011).
5. Bates, K.E. *et al.* Pediatric cardiovascular safety: challenges in drug and device development and clinical application. *Am. Heart. J.* **164**, 481–492 (2012).
6. Sammons, H. Ethical issues of clinical trials in children: a European perspective. *Arch. Dis. Child.* **94**, 474–477 (2009).
7. EMA. Guideline on clinical trials in small populations (2006).
8. Baiardi, P., Giaquinto, C., Girotto, S., Manfredi, C. & Ceci, A. Innovative study design for paediatric clinical trials. *Eur. J. Clin. Pharmacol.* **67**(suppl. 1), 109–115 (2011).
9. Smania, G., Baiardi, P., Cella, M. & Magni, P. Model-based comparison of alternative study designs in pediatric trials. Part II: Bayesian approaches. *CPT Pharmacometrics Syst. Pharmacol.* (submitted).
10. Wellek, S. & Blettner, M. On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications. *Dtsch. Arztebl. Int.* **109**, 276–281 (2012).
11. Ruperto, N. *et al.* Abatacept in children with juvenile idiopathic arthritis: a randomised, double-blind, placebo-controlled withdrawal trial. *Lancet* **372**, 383–591 (2008).
12. Lovell, D.J. *et al.* Adalimumab with or without methotrexate in juvenile rheumatoid arthritis. *N. Engl. J. Med.* **359**, 810–820 (2008).
13. Todd, S., Whitehead, A., Stallard, N. & Whitehead, J. Interim analyses and sequential designs in phase III studies. *Br. J. Clin. Pharmacol.* **51**, 394–399 (2001).
14. Sébille, V. & Bellissant, E. Sequential methods and group sequential designs for comparative clinical trials. *Fundam. Clin. Pharmacol.* **17**, 505–516 (2003).
15. 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).
16. Nehus, E.J., Mouksassi, S., Vinks, A.A. & Goldstein, S. Meropenem in children receiving continuous renal replacement therapy: clinical trial simulations using realistic covariates. *J. Clin. Pharmacol.* **54**, 1421–1428 (2014).
17. Piana, C. *et al.* A model-based approach for the evaluation of once daily dosing of lamivudine in HIV-infected children. *Br. J. Clin. Pharmacol.* **77**, 852–860 (2014).
18. Willmann, S. *et al.* Development of a paediatric population-based model of the pharmacokinetics of rivaroxaban. *Clin. Pharmacokinet.* **53**, 89–102 (2014).
19. Cella, M., Danhof, M. & Della Pasqua, O. Adaptive trials in paediatric development: dealing with heterogeneity and uncertainty in pharmacokinetic differences in children. *Br. J. Clin. Pharmacol.* **74**,346–353 (2012).
20. Brochot, A. *et al.* Model-based once-daily darunavir/ritonavir dosing recommendations in pediatric HIV-1-infected patients aged ≥3 to <12 years. *CPT Pharmacometrics Syst. Pharmacol.* **4**, 406–414 (2015).
21. Jadhav, P.R., Zhang, J.& Gobburu, J.V. Leveraging prior quantitative knowledge in guiding pediatric drug development: a case study. *Pharm. Stat.* **8**, 216–224 (2009).
22. Salazar, D.E. *et al.* The use of modeling and simulation to guide clinical development of olmesartan medoxomil in pediatric subjects. *Clin. Pharmacol. Ther.* **91**, 250–256 (2012).
23. Reif, S., Schultze-Mosgau, M. & Sutter, G. From adults to children: simulation-based choice of an appropriate sparse-sampling schedule. *Clin. Pharmacol. Ther.* **14**, 189–200 (2012).
24. Cella, M., Danhof, M. & Della Pasqua, O. Bridging strategies for drug combinations in pediatric indications. *Clin. Pharmacol. Ther.* **91**, 726–733 (2012).
25. FDA. Topomax Label (2012).
26. EMA. Topomax package leaflet (2009).
27. French, J.A., Cabrera, J., Emir, B., Whalen, E. & Lu, F. Designing a new proof-of-principle trial for treatment of partial seizures to demonstrate efficacy with minimal sample size and duration-a case study. *Epilepsy Res.* **106**, 230–236 (2013).
28. Eriksson, A.S., Nergardh, A. & Hoppu, K. The efficacy of lamotrigine in children and adolescents with refractory generalized epilepsy: a randomized, double-blind, crossover study. *Epilepsia* **39**, 495–501. (1998).
29. O'Brien, P.C. & Fleming, T.R. A multiple testing procedure for clinical trials. *Biometrics* **35**, 549–556 (1979).
30. Pocock, S.J. Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64**, 191–199 (1977).
31. van der Lee, J.H., Wesseling, J., Tanck, M.W. & Offringa, M. Sequential design with boundaries approach in pediatric intervention research reduces sample size. *J. Clin. Epidemiol.* **63**, 19–27 (2010).
32. Whitehead, J. *The Design and Analysis of Sequential Clinical Trials*, revised 2nd edn. (John Wiley & Sons, Chichester, UK, 1997).
33. WHO. Weight-for-age percentiles. <www.who.int> (2006).
34. Beal, S., Sheiner, L.B., Boeckmann, A. & Bauer, R.J. NONMEM 7.2.0 User's Guides (1989–2011). Icon Development Solutions, Ellicott City, MD.
35. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing (2013).
36. Bower, P. *et al.* Interventions to improve recruitment andretention in clinical trials: a survey and workshop to assess current practice and future priorities. *Trials.* **15**:399 (2014).
37. Sebille, V. & Bellissant, E. Comparison of four sequential methods allowing for early stopping of comparative clinical trials. *Clin. Sci.* **98**, 569–578 (2000).
38. Emerson, S.S., Kittelson, J.M. & Gillen, D.L. Frequentist evaluation of group sequential clinical trial designs. *Stat. Med.* **26**, 5047–5080 (2007).
39. Perucca, E. What clinical trial designs have been used to test antiepileptic drugs and do we need to change them? *Epileptic Disord.* **14**, 124–131 (2012).
40. EMA. Concept paper on extrapolation of efficacy and safety in medicine development (2012).
41. Anderson, B.J. & Holford, N.H. Understanding dosing: children are small adults, neonates are immature children. *Arch. Dis. Child.* **98**, 737–744 (2013).
42. Hlavin, G., Koenig, F., Male, C., Posch, M. & Bauer, P. Evidence, eminence and extrapolation. *Stat. Med.* (in press) (2016).

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