Article

Handling Poor Accrual in Pediatric Trials: A Simulation Study Using a Bayesian Approach

Danila Azzolina [1,2] , Giulia Lorenzoni [1] , Silvia Bressan [3] , Liviana Da Dalt [3], Ileana Baldi [1] and Dario Gregori [1,*]

1   Unit of Biostatistics, Epidemiology and Public Health, Department of Cardiac Thoracic Vascular Sciences and Public Health, University of Padova, 35128 Padova, Italy; danila.azzolina@uniupo.it (D.A.); giulia.lorenzoni@unipd.it (G.L.); ileana.baldi@unipd.it (I.B.)
2   Department of Translational Medicine, University of Eastern Piedmont, 28100 Novara, Italy
3   Department of Women's and Children's Health, University of Padova, 35128 Padova, Italy; silvia.bressan.1@unipd.it (S.B.); liviana.dadalt@unipd.it (L.D.D.)
*   Correspondence: dario.gregori@unipd.it

Abstract: In the conduction of trials, a common situation is related to potential difficulties in recruiting the planned sample size as provided by the study design. A Bayesian analysis of such trials might provide a framework to combine prior evidence with current evidence, and it is an accepted approach by regulatory agencies. However, especially for small trials, the Bayesian inference may be severely conditioned by the prior choices. The Renal Scarring Urinary Infection (RESCUE) trial, a pediatric trial that was a candidate for early termination due to underrecruitment, served as a motivating example to investigate the effects of the prior choices on small trial inference. The trial outcomes were simulated by assuming 50 scenarios combining different sample sizes and true absolute risk reduction (ARR). The simulated data were analyzed via the Bayesian approach using 0%, 50%, and 100% discounting factors on the beta power prior. An informative inference (0% discounting) on small samples could generate data-insensitive results. Instead, the 50% discounting factor ensured that the probability of confirming the trial outcome was higher than 80%, but only for an ARR higher than 0.17. A suitable option to maintain data relevant to the trial inference is to define a discounting factor based on the prior parameters. Nevertheless, a sensitivity analysis of the prior choices is highly recommended.

Academic Editor: Paul B. Tchounwou

1. Introduction

Difficulties in the enrolment of the overall trial sample size, as indicated at the design stage, could be caused by several factors (i.e., high costs, regulatory barriers, narrow eligibility criteria, and cultural attitudes toward research in almost all research fields). Effects can be different depending on the population's characteristics and the intervention under evaluation [1].

Prior research evaluating the reasons for termination across a broad range of trials reported that insufficient enrolment is the most common reason, with a frequency ranging from 33.7% to 57%, depending on the definition used [2,3]. The slow or low accrual problem is common in clinical research on adults, primarily in oncology [4–6] and cardiology [7], as well as in pediatric research, in which 37% of clinical trials are terminated early due to inadequate accrual [8]. Pediatrics is a research field that requires particular attention, since accrual issues are associated with methodological and ethical challenges [9]. It is essential to consider that the management and conduct of pediatric trials are more complicated than those of adult trials in terms of practical, ethical, and methodological problems [10].

From a statistical point of view, low accrual results in a reduced sample size, compromising the ability to accurately answer the primary research question due to a reduction in

the likelihood of detecting a treatment effect [11]. The scientific community has conveyed that early termination of a trial due to poor accrual leads to inefficiency in clinical research, with consequent increases in costs [12] and a waste of resources, as well as a waste of the efforts of the children involved in the trial [13].

For these reasons, alternative and innovative approaches to pediatric clinical trial design have been a recent topic of debate in the scientific community [9,14]. Alternative methods for pediatric trial design and analysis have been proposed by recent guidelines in the field, i.e., the ICH (International Council for Harmonisation) Topic E11 guidelines [15], the guidance for trial planning and design in the pediatric context [16], and the EMA (European Medicines Agency) guidelines [16–18].

It is noteworthy that data from trials terminated prematurely for poor accrual can provide useful information for reducing the uncertainty about the treatment effect in a Bayesian framework [11].

In recent years, Bayesian methods have increasingly been used in the design, monitoring, and analysis of clinical trials due to their flexibility [19,20]. Considering the research setting described in this work, the Bayesian methods used for accrual monitoring are also interesting [21]. These methods are well suited to designing and analyzing studies conducted with small sample sizes and are particularly appropriate for studies involving children, even in cases of rare disease outcomes [9].

In clinical trials that are candidates for early termination due to poor accrual reasons, a Bayesian approach may be useful for incorporating the available knowledge on the investigated treatment effect, reported in the literature or elicited by experts' opinions [22]. In addition, in a Bayesian setting, prior information combined with data may support the final inference for a trial conducted on a limited number of enrolled patients [23,24].

In pediatric trials, for example, the awareness that a treatment is effective in adults increases the probability of its efficacy in children. This knowledge may be quantitatively translated into a prior probability distribution [9,14].

However, when there is a small sample size, the final inference may be severely conditioned by a misleading prior definition [24]. In this framework, the Food and Drug Administration (FDA) suggests performing a sensitivity analysis on prior definitions [25], especially for very small sample sizes [26]. In this regard, the power prior approach is used to design and analyze small trials to control for the weight of historical information, translated into prior distributions, through prior discounting factors [27,28]. The use of historical information to define the prior distribution in a nonparametric context is a method recently used in the literature [29]. Informative prior elicitation is typically a challenging task even in the presence of historical data (objective prior) [30]. Ibrahim and Chen [28] proposed the power prior approach to incorporate the historical data in the analysis of a current study. The method is based on the raising of the likelihood function of the historical data to a power parameter between 0 and 1 (power parameter). This parameter represents the proportion of the historical data incorporated in the prior.

Hobbs modified the conventional approach, accounting for commensurability of the information in the historical and current data to determine how much the historical information is used in the inference [31]. Other power-prior proposals calibrate the type I error by controlling the degree of similarity between the new and historical data [32,33]. The prior-data conflict has also been addressed and incorporated in the power prior in a commensurability parameter defined by using a measure of distribution distance in a group sequential design clinical trial [34]. A mixture of priors, for the one-parameter exponential family, has been also considered in a sequential trial, to incorporate the historical data accounting for rapid reaction to prior-data conflicts by adding an extra weakly-informative mixture component [35].

In general, the power prior approach is widely used for the design and analysis of clinical trial data. The method is useful for handling problems related to a lack of exchangeability between the historical and current data, and the risk that prior information overwhelms the clinical trial data information [27].

The optimal amount of discounting factors for an informative prior remains to be discussed [14].

This study investigated the effects of the prior choices on the final inference, especially for studies conducted with limited sample sizes, such as pediatric trials. A pediatric trial candidate for early termination due to underrecruitment, the RESCUE trial, served as a motivating example for the simulation study proposed.

A set of possible trial outcomes were simulated. The simulation plan was designed to evaluate the effects of the prior choices on the trial results by evaluating different scenarios depending on the number of patients involved in the study and the magnitude of the true treatment effect.

2. Materials and Methods

2.1. Motivating Example

The RESCUE trial was a randomized controlled double-blind trial. The purpose of the study was to evaluate the effect of adjunctive oral steroids in preventing renal scarring in young children and infants with febrile urinary tract infections. The primary outcome was the renal scar absolute risk reduction (ARR) between the treatment arms. The study was designed expecting an ARR of 0.20 to determine a renal scar reduction from 40% to 20%.

After two years, only 17 recruited patients completed the follow-up for the study outcome (6 in treatment and 11 in control) due to procedural problems and poor compliance with the study therapy and final diagnosis [16–18].

2.2. Simulation Plan

The possible trial outcomes were simulated by assuming several scenarios combining different sample sizes and true ARRs. The simulated data were analyzed via the Bayesian approach using a beta prior distribution whose parameters were derived from trials conducted in research settings similar to the RESCUE trial. The beta-binomial model was considered because it is the most widely used approach among the Bayesian methods to summarize event rates in clinical trials [36]. This parametrization is easily computationally tractable and is very precise [37].

Informative, low-informative, and uninformative priors were selected for the analyses according to the discounting levels placed on the prior parameters.

The classical, non-Bayesian approach was considered a benchmark.

This simulation study is defined by:

1. Data generation hypotheses.
2. Analysis of simulated data.
3. Presentation of the results of simulations.

A flowchart synthesizing the simulation plan is reported in Figure S1, Supplementary Material.

2.3. Data Generation Hypotheses

2.3.1. Simulation Scenarios

The simulation plan consisted of 50 scenarios. Each scenario represents a single combination of the treatment effect (ARR) and the sample size used to generate the data. Fifty scenarios were considered, since they combined ten different sample sizes (ranging from 15 to 240) within five assumed ARRs (Table 1). The ARR ranged from −0.07 to −0.27, with an increment of 0.07, according to the treatment effects suggested by the literature [38,39].

Table 1. Simulation scenarios.

Scenario	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Sample size	15	40	65	90	115	140	165	190	215	240	15	40	65	90	115	140	165	190	215	240	15	40	65	90	115
True ARR	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.17	0.17	0.17	0.17	0.17
Scenario	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**
Sample size	140	165	190	215	240	15	40	65	90	115	140	165	190	215	240	15	40	65	90	115	140	165	190	215	240
True ARR	0.17	0.17	0.17	0.17	0.17	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27

ARR=Absolute Risk Reduction.

2.3.2. Data Generation within Scenarios

For each scenario, the trial data were randomly generated 5000 times. The data were drawn from a binomial random variable, assuming a true event rate in the control arm of $\pi_{control} = 0.33$. This event rate is in-between the results provided by Huang et al. [38] and Shaikh et al. [39] for the control group.

The treatment arm data were generated using a binomial random variable hypothesizing an ARR, one for each experiment, in compliance with the simulation plan provided in Table 1, where the sample size is showed overall. However, it is assumed that the control arm contains 60% of the sample size to reflect the group imbalance in the motivating example.

2.4. Analysis of the Simulated Data

The 5000 randomly generated data points were analyzed via the Bayesian method by considering: (1) the informative prior, (2) the low-informative prior, and (3) the uninformative prior. A frequentist analysis was performed for comparison purposes.

The data were simulated 5000 times by a binomial random variable in a frequentist approach. For each of the repeated simulations, the ARR was calculated and the binomial confidence interval was estimated.

2.4.1. Prior Definition

A mixture of beta priors was considered for the outcome evaluation, using data provided by the literature [38,39]. The clinical trial results were combined in a mixture of distributions. The beta distributions comprising the mixture of priors for each scar event rate in the treatment and control groups were derived from other trials' historical information [27].

The functional form of the distribution is characterized by the shape α and scale β parameters $\Pi \sim \text{Beta}(\alpha, \beta)$ [40], where Π is the parameter that characterizes the event rate on which to make inference. The shape value α is defined by the number of events x observed in other trials, while the β value corresponds to the number of subjects not experiencing the event $(n - x)$ [41].

1. Huang et al. [38] reported probabilities of scarring of $\hat{\pi}_{\text{treat (Huang)}} = \frac{6}{18} = 0.33$ and $\hat{\pi}_{\text{control (Huang)}} = \frac{39}{65} = 0.66$ in the treatment and control arms, respectively. Considering this information, the informative beta prior can be derived as:

$$\Pi_{\text{treat (Huang)}} \sim \text{Beta}(6, 12)$$

$$\Pi_{\text{control (Huang)}} \sim \text{Beta}(39, 26)$$

2. Shaikh et al. [39] reported, instead, probabilities of scarring of $\hat{\pi}_{\text{treat (Shaikh)}} = 0.098$ (12|123) and $\hat{\pi}_{\text{control (Shaikh)}} = 0.168$ (22|131) in the treatment and control arms,

respectively. Considering this information, the informative beta prior can be derived as:

$$\Pi_{\text{treat (Shaikh)}} \sim \text{Beta}(12, 111)$$

$$\Pi_{\text{control(Shaikh)}} \sim \text{Beta}(22, 109)$$

The information was combined in a mixture of beta priors:

- For the treatment arm, the beta mixture is defined as:

$$\Pi_{\text{treat}} = \gamma \Pi_{\text{treat (Huang)}} + (1 - \gamma)\Pi_{\text{treat (Shaikh)}}$$

The expected value for the mixture random variable is, for the treatment arm, a weighted mean of the expectations over the mixture components:

$$\text{E}\left[\Pi_{\text{treat}}\right] = \gamma \text{E}[\Pi_{\text{treat(Huang)}}] + (1 - \gamma)\text{E}[\Pi_{\text{treat(Shaikh)}}]$$

If we denote the beta shape $\alpha_{\text{treat(Huang)}}$ and $\alpha_{\text{treat(Shaikh)}}$, respectively for the Huang and Shaikh studies, and $\beta_{\text{treat(Huang)}}$ and $\beta_{\text{treat(Shaikh)}}$ the scales for the considered studies, the mixture expected value may be computed as:

$$\text{E}[\Pi_{\text{treat}}] = \gamma \text{E}\left[\Pi_{\text{treat(Huang)}}\right] + (1 - \gamma)\text{E}\left[\Pi_{\text{treat(Shaikh)}}\right]$$

$$\text{E}[\Pi_{\text{treat}}] = \gamma \frac{\alpha_{\text{treat(Huang)}}}{\alpha_{\text{treat(Huang)}} + \beta_{\text{treat(Huang)}}} + (1 - \gamma)\frac{\alpha_{\text{treat(Shaikh)}}}{\alpha_{\text{treat(Shaikh)}} + \beta_{\text{treat(Shaikh)}}}$$

$$= \gamma \frac{6}{6 + 12} + (1 - \gamma)\frac{12}{12 + 111}$$

If we assume an equal weight value $\gamma = 0.5$, $\text{E}[\Pi_{\text{treat}}] = 0.215$.

- The mixture variance is given by:

$$\text{Var}[\Pi_{\text{treat}}] = \left[\gamma\left(\text{Var}\left[\Pi_{\text{treat(Huang)}}\right] + \text{E}\left[\Pi_{\text{treat(Huang)}}\right] - \text{E}[\Pi_{\text{treat}}]\right)\right] +$$
$$+ \left[(1 - \gamma)\left(\text{Var}\left[\Pi_{\text{treat(Shaikh)}}\right] + \text{E}\left[\Pi_{\text{treat(Shaikh)}}\right] - \text{E}[\Pi_{\text{treat}}]\right)\right]$$

where the variances of the mixture components are:

$$\text{Var}\left[\Pi_{\text{treat(Huang)}}\right] = \frac{\alpha_{\text{treat(Huang)}}\beta_{\text{treat(Huang)}}}{(\alpha_{\text{treat(Huang)}} + \beta_{\text{treat(Huang)}})^2(\alpha_{\text{treat(Huang)}} + \beta_{\text{treat(Huang)}} + 1)}$$

$$\text{Var}[\Pi_{\text{treat(Shaikh)}}] = \frac{\alpha_{\text{treat(Shaikh)}}\beta_{\text{treat(Shaikh)}}}{(\alpha_{\text{treat(Shaikh)}} + \beta_{\text{treat(Shaikh)}})^2(\alpha_{\text{treat(Huang)}} + \beta_{\text{treat(Shaikh)}} + 1)}$$

Equal weight was assumed for the components of the mixture, therefore, $\gamma = 0.5$, $\text{E}[\Pi_{\text{treat}}] = 0.215$, and $\text{SD}[\Pi_{\text{treat}}] = 0.08$.

- For the treatment arm, the mixture is defined as:

$$\Pi_{\text{control}} = \gamma \Pi_{\text{control (Huang)}} + (1 - \gamma)\Pi_{\text{control (Huang)}}$$

with $\text{E}[\Pi_{\text{control}}] = 0.38$ and $\text{SD}[\Pi_{\text{control}}] = 0.05$ and $\gamma = 0.5$.

2.4.2. Discounting the Priors: The Power Prior Approach

Different levels of penalization (discounting) were provided for the historical information using a power prior approach [28] to perform a sensitivity analysis on the prior choices. The historical information can be included in the final inference using a Beta(α_1, β_1) prior, where:

$$\alpha_1 = 1 + \alpha_0 d_0$$

$$\beta_1 = 1 + \beta_0 d_0$$

The α_0 and β_0 values are the parameters defined by the number of successes and failures derived from the literature and are α_0 and β_0, respectively. The value d_0 defines the amount of historical information to be included in the final inference. The discounting factor is otherwise defined as $(1 - d_0) \times 100$ and represents the level of penalization (discounting) of the historical information derived from other studies.

1. If $d_0 = 0$, the data provided by the literature are not considered, indicating a 100% discount of the historical information. According to this scenario, the prior is an uninformative $\text{Beta}(1,1)$ distribution.
2. If $d_0 = 1$, then all of the information provided by the literature is considered in setting the prior, indicating a 0% discount of the historical data.

Analyses of the simulated trials were conducted using three different priors:

-   Power prior without discounting (informative, $d_0 = 1$). A beta informative prior was derived considering the number of successes and failures found in the literature [42], as defined in the method section.
-   Power prior 50% discounting (low-informative, $d_0 = 0.5$). The beta prior with a 50% discount, defined in the literature as a substantial-moderate discounting factor [43], was defined based on the beta parameters comprising the mixture of priors specified in the informative scenario.
-   Power prior 100% discounting (uninformative, $d_0 = 0$). A mixture of $\text{Beta}(1,1)$ priors was defined.

Effective Sample Size (ESS) Calculation

The ESS was computed on the mixture of beta distribution by using the Morita approach to quantify the prior influence on the final inference using the RBesT package in R (R Foundation for Statistical Computing, Vienna, Austria) [44]. For the mixture of beta prior (equal weight) without power prior discounting ($d_0 = 1$), an ESS of 55 and 98 was achieved for treatment and control arm. However, discounting the beta parameters for $d_0 = 0.5$ (low-informative prior), the ESS is equal to 24 and 48.

The prior distributions are presented in Figure 1.

2.4.3. Posterior Estimation

A beta-binomial model was employed to analyze the difference in event rates between arms [45]. The posterior distribution for the ARR outcome requires the estimation of the posterior distribution of the scar proportion in each arm separately, and was computed with the following Markov chain Monte Carlo (MCMC) resampling procedure [46]:

-   A first resampling of the proportion of scarring Π^*_{treat} from $\Pi_{\text{treat}}|X_{\text{treat}}$, which is the posterior distribution for the treatment group.
-   A second resampling of Π^*_{control} from $\Pi_{\text{control}}|X_2$.
-   The posterior distribution for the parameter related to the difference in proportions was obtained by calculating $\text{ARR} = \Pi^*_{\text{treat}} - \Pi^*_{\text{control}}$ from the distributions previously resampled [47].

The resampling procedures were performed using an MCMC estimation algorithm, as indicated in the literature [46], using 3 chains, 6000 iterations, and 1000 adaptations.

An example of the inference results is reported in the Supplementary Material, showing the priors and the posterior distributions calculated on a single database generated by assuming an ARR equal to 0.17.

The computations were performed using OpenBUGS (Free Software Foundation, Boston, MA, USA) [48] and R version 3.3.2 [49]; the simulation R codes are reported in the Supplementary Material.

Int. J. Environ. Res. Public Health **2021**, *18*, 2095

7 of 16
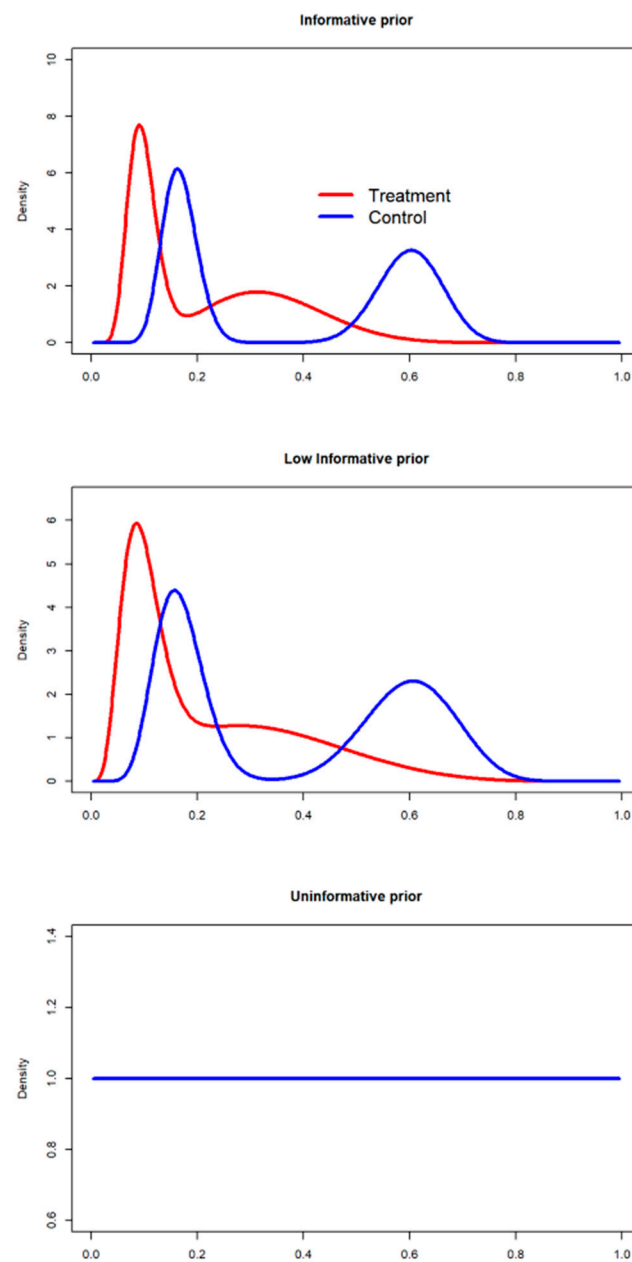
Figure 1. Prior distributions: The prior distributions are defined by an equal-weighted mixture ($\gamma = 0.5$) of beta priors. The components of the mixture prior are, for the treatment arm, $\Pi_{\text{treat (Huang)}} \sim \text{Beta}(6, 12)$ and $\Pi_{\text{treat (Shaikh)}} \sim \text{Beta}(12, 111)$. The mixture of priors ($\gamma = 0.5$) for the control arm is defined by $\Pi_{\text{control (Huang)}} \sim \text{Beta}(39, 26)$ and $\Pi_{\text{control(Shaikh)}} \sim \text{Beta}(22, 109)$. No discounting on the beta priors parameters has been provided ($d_0 = 1$) for the Informative priors. The information has been partially discounted for the low-informative prior scenario ($d_0 = 0.5$). The priors parameters are full discounted for the uninformative prior scenario ($d_0 = 0$), collapsing to a $\text{Beta}(1, 1)$ distribution.

2.4.4. Convergence Assessment

The Geweke method [50] was considered to assess the convergence of the MCMC results within iterations. Geweke's statistics test was computed for each analysis conducted on the simulated data. Geweke's Z-score plot was also visually inspected to assess the convergence.

2.5. Results of the Simulations

Four sets of 200 results summarizing 50 scenarios in combination with four methods of analysis were defined as:

1. The proportion of the 5000 simulated trials for which the credibility intervals (CIs), or confidence intervals, for a frequentist analysis do not contain an ARR equal to 0. The proportion of intervals not containing the 0 and containing the data generator ARR was also calculated.
2. The mean length across 5000 simulated trials of the CI.
3. The mean of the posterior median estimate across 5000 simulated trials or the mean of the point-estimated ARR across 5000 simulated trials for the frequentist analysis.
4. The mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\text{ARR}_{\text{true}} - \text{A}\hat{\text{R}}\text{R}_t}{\text{ARR}_{\text{true}}}\right|$$

ARR_{true} is the true treatment effect considered to generate the data; $\text{A}\hat{\text{R}}\text{R}_t$ is the estimated treatment effect (posterior median, or point estimate, for the frequentist analysis) achieved for each simulation t within the n = 5000 simulated trials.

3. Results

The proportion of 5000 simulated trials ensuring that the 95% CI does not contain an ARR equal to zero is greater than 90% for all of the informative scenarios, even if the sample size is smaller than 50, except for the 0.07 true ARR. For the 0.07 ARR, this proportion declines as the data used to estimate the likelihood increases (Figure 2, Panel A). This proportion is higher than 80% only for sample sizes greater than 70, and the true ARR is greater than 0.17 for the low-informative priors (Figure 2, Panel B). The pattern of the simulation results is similar, considering the proportion of simulations for which the CI does not include the 0, and includes the true data generator ARR (Figure S3, Supplementary Material).

Similar behavior is observed among the uninformative Bayesian (Figure 2, Panel C) and frequentist (Figure 2, Panel D) estimates, for which this proportion reaches 80% for an ARR greater than 0.22 and sample sizes greater than 120.

The 95% CI length decreases as the sample size increases for all of the Bayesian parametrizations and the frequentist estimates (Table S1, Supplementary Material). The informative (Figure 3, Panel A) and low-informative priors (Figure 3, Panel B) showed more variability in the posterior length of the CIs across different true ARR values. The CI lengths are more similar for different data generation ARR assumptions for the uninformative (Figure 3, Panel C) and frequentist (Figure 3, Panel D) simulations. In general, especially for smaller sample sizes, the estimates are less precise for the frequentist and Bayesian uninformative prior scenarios than for the informative and low-informative prior estimates (Table 1).

The posterior median ARR estimates are influenced by the prior choices, especially for the informative prior. The estimated ARRs are similar to each other for smaller sample sizes across the true treatment effect, while the posterior median ARR estimates converge to the true ARR for larger sample sizes (Figure 4, Panel A). A similar pattern is observed for the low-informative scenarios; however, for smaller sample sizes, greater variability in the posterior median estimates is observed across the different ARRs used to generate the data (Figure 4, Panel B). The ARR is overestimated for small sample sizes in the uninformative prior scenarios (Figure 4, Panel C). Instead, the frequentist estimates across the simulated trial are similar to the true treatment effect for all of the sample sizes (Figure 4, Panel D).

The MAPE estimate decreases as the sample size increases for all the prior parametrizations (Table S1, Supplementary Material). A lower true ARR (i.e., 0.07) ensures a decreasing effect that is more evident than a higher true ARR (Table S1, Supplementary Material).

Int. J. Environ. Res. Public Health **2021**, *18*, 2095

9 of 16

Also, the MAPE seems to be constant for a higher true ARR in informative (Figure 5, Panel A) and low-informative prior (Figure 5, Panel B) simulations. For the uninformative (Figure 5, Panel C) and frequentist scenarios (Figure 5, Panel D), instead, a reduction in MAPE is also evident for higher true ARR values. The MAPE values are higher for the frequentist scenarios than all of the Bayesian estimates, including those provided via the uninformative prior (Table S1, Supplementary Material).
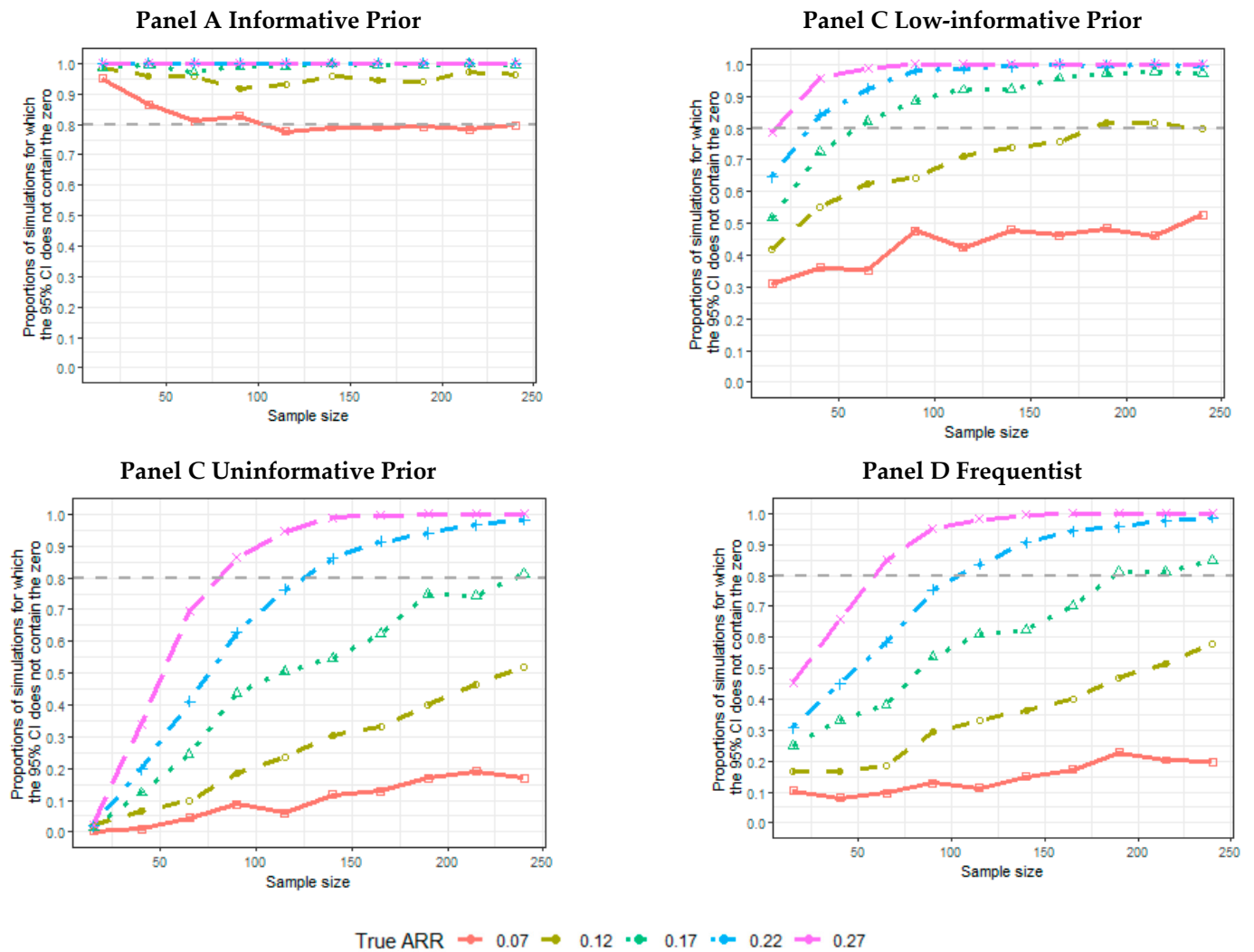

Figure 2. The proportion of CIs within simulated trials not including the zero absolute risk reduction (ARR) according to the sample size, and true ARR for informative prior (Panel A), low-informative prior (Panel B), uninformative prior (Panel C), and frequentist analysis (Panel D).

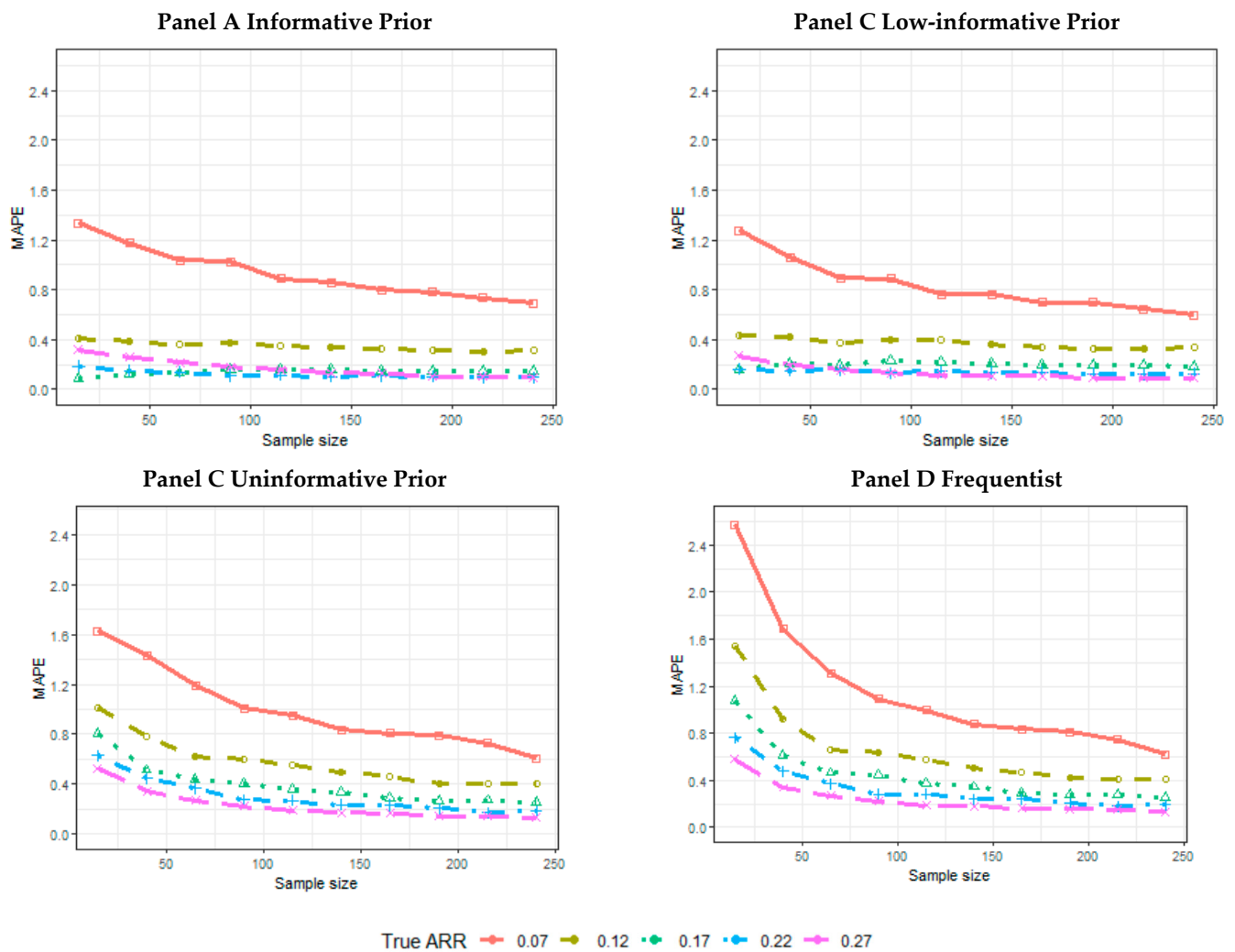*Int. J. Environ. Res. Public Health* **2021**, *18*, 2095

10 of 16
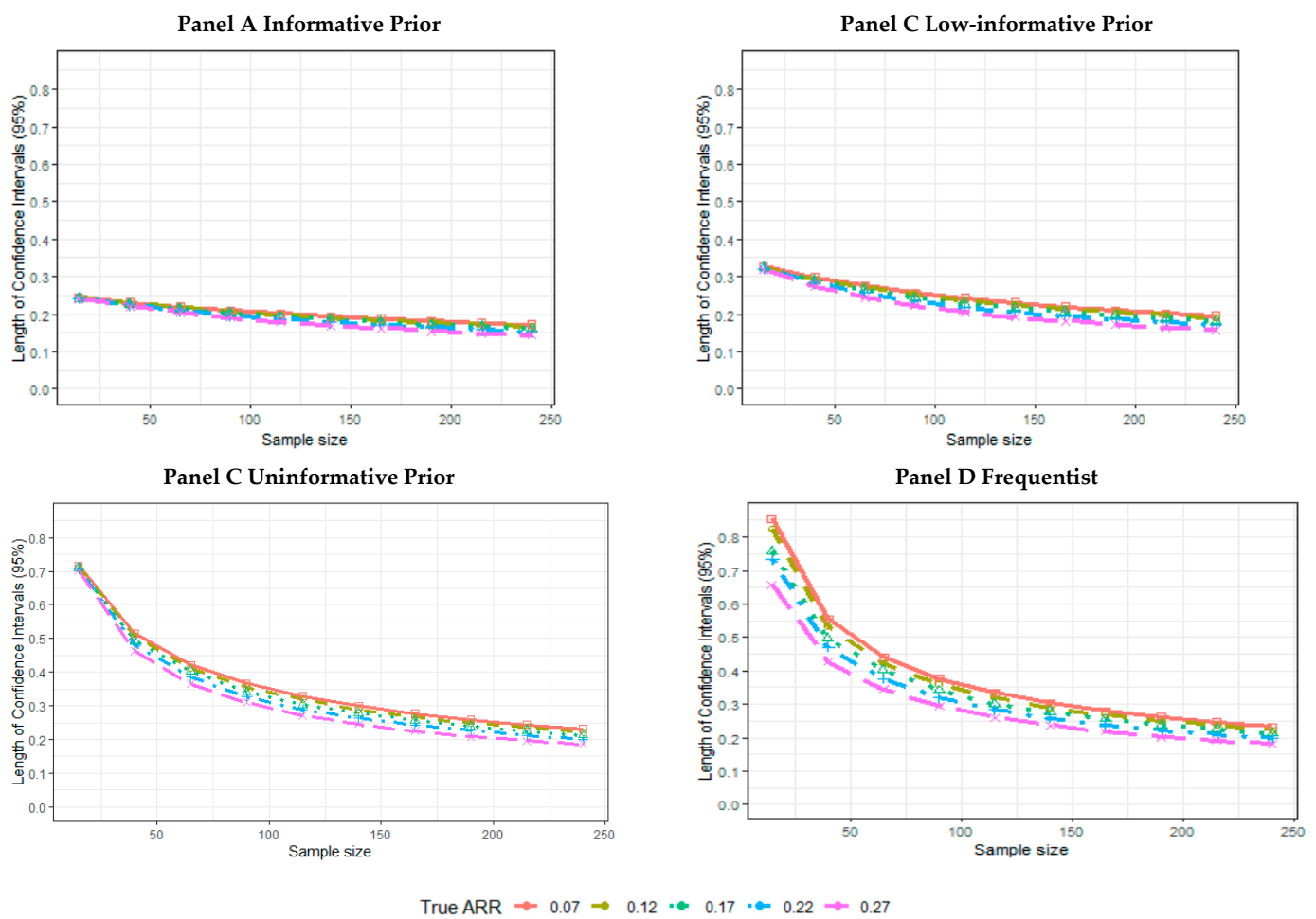
Figure 3. Simulation results for the 95% CI length according to the sample size and true ARR for informative prior (Panel A), low-informative prior (Panel B), uninformative prior (Panel C), and frequentist analyses (Panel D).

The hypothesis of the stationarity of the chain was not rejected according to Geweke's statistic for all of the analyses conducted on the simulated data and for all of the prior parametrizations. The Z-scores within iterations was also visually inspected. An example within simulations (ARR = 0.07 and sample size = 65) is reported in the Figure S2, Supplementary Material. The Z-score lies within the acceptance stationarity region (± 2) or all chains and all the prior parametrizations; the pattern is very similar for all the considered scenarios.

An example of a possible inference result is shown in the Supplementary Material. The posteriors were calculated for a generated trial data reporting 8 events over 56 in the treatment arm ($\hat{\pi}_{treat} = 0.14$) and 30 events over 84 in the control arm ($\hat{\pi}_{control} = 0.36$). The data generator ARR is 0.17, while the observed ARR is 0.22. Considering the different priors, the inference results are located in mean on the same event rate; however, the uncertainty in the posterior distribution increases, considering the uninformative prior assumption (Figure S4, Supplementary Material).

Int. J. Environ. Res. Public Health **2021**, *18*, 2095

11 of 16
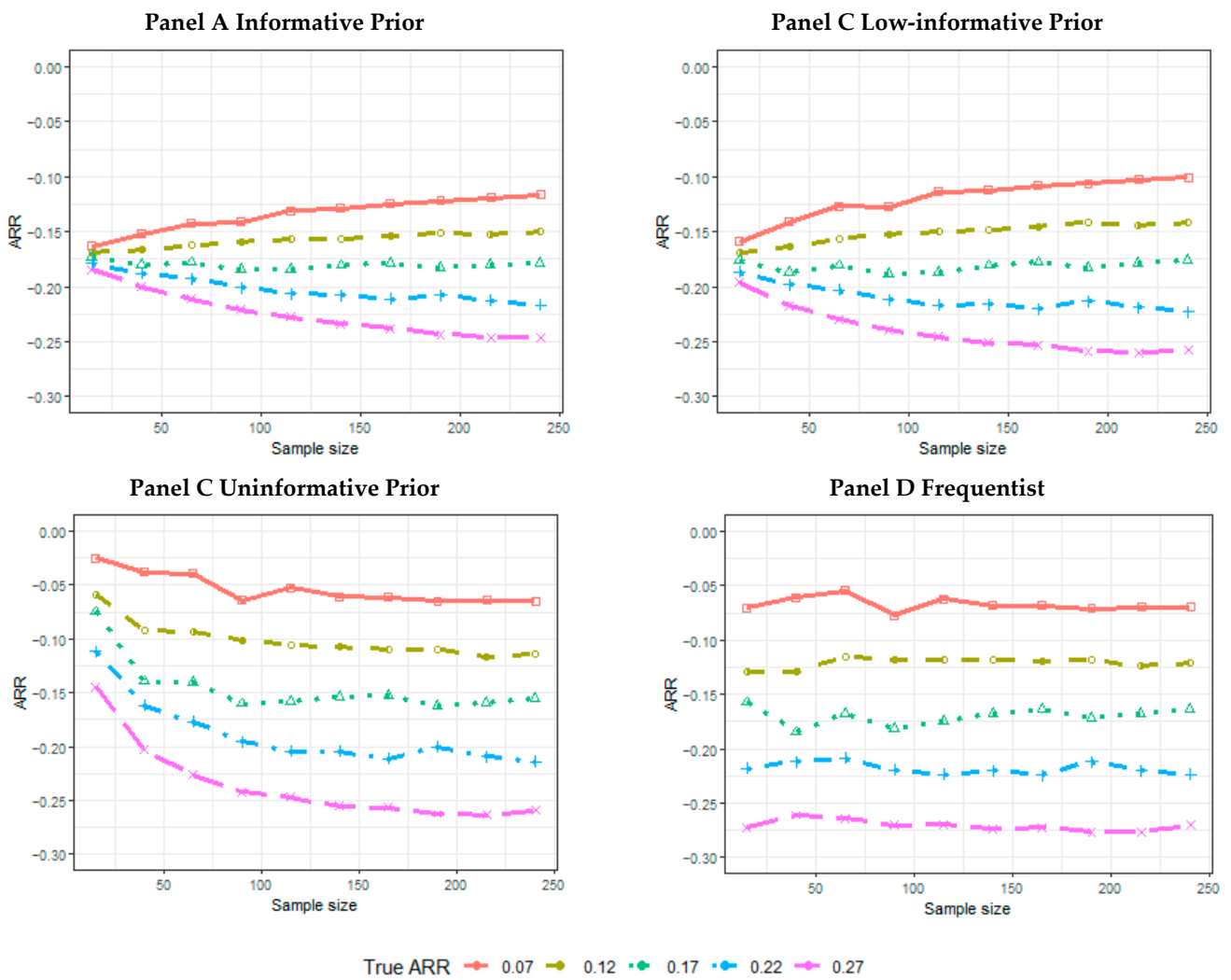
Figure 4. Simulation results for the estimated ARR (posterior median, or point estimate, for frequentist analysis) according to the sample size and true ARR for informative prior (Panel A), low-informative prior (Panel B), uninformative prior (Panel C), and frequentist analyses (Panel D).

Figure 5. Simulation results for the mean absolute percentage error (MAPE) estimate according to the sample size and true ARR for informative prior (Panel A), low-informative prior (Panel B), uninformative prior (Panel C), and frequentist analyses (Panel D).

4. Discussion

Regulatory agencies advocate an increase in pediatric research, which is motivated by the need for more information on treatment labeling to guide pediatricians and to offer more suitable and safe treatments for children [14]. However, in various cases, pediatric trials have demonstrated difficulties in enrolling participants [51]. The RESCUE trial represents a typical example of a complex trial in pediatric research affected by poor accrual. The difficulties encountered in the enrolment and retention of participants are related to procedural problems related to the study protocol [51,52] and poor adherence to the therapy.

Bayesian data analysis may overcome challenges in the conduction of trials similar to the RESCUE study, allowing investigators to combine information provided by current trial data with evidence provided by the literature, as recommended by regulatory agencies to deal with small sample sizes [15].

The present findings show that Bayesian inference can detect a small treatment effect for small sample sizes (lower than 50), even if the prior is fully uninformative compared to a maximum likelihood approach. This result confirms the potential benefits of using a Bayesian method on small sample sizes. However, the literature suggests paying attention to the use of uninformative prior distributions for small clinical trials, because there is the possibility of including in the final inference extreme treatment effects that are potentially

Int. J. Environ. Res. Public Health **2021**, *18*, 2095

13 of 16

unexpected from a clinical point of view. For this reason, it is suggested to use evidence from previous trials to inform these prior distributions [53].

For this reason, a key issue in Bayesian analysis is the choice of prior. This simulation study demonstrated that, especially for small studies, the trial results could be influenced by the prior choices and weakly influenced by the data when using fully informative priors. In particular, a prior distribution incorporating favorable treatment effect information on small sample sizes is likely to conditionate the inference in favor of the treatment, even if, in truth, the effect is null or minimal. All of this implies that the prior in these contexts should be defined by using validated empirical evidence [27]. Conversely, this study suggests that the full informative prior elicited by considering large effect size tends to direct the inference towards the existence of a treatment effect for all the sample size scenarios. For this reason, we recommend, especially when the treatment effect hypothesized for the study design is large and the sample size is small, the use of a low-informative prior for achieving more data-driven results.

The situation is different if a discounting factor is placed on the prior parameters. Looking at the estimated values of ESS, the historical information retained in the prior in the low-informative scenario is halved, compared to the informative parameterization. This implies that the inference is more data-oriented, assuming a discounting of 0.5. The probability of confirming the trial results is demonstrated to be more data-dependent and, for sample sizes less than 50, is higher than 80% only for ARRs higher than 0.17.

As the power prior parameter increases, the prior becomes more informative, and the estimated precision (length of CI) increases. Looking at the differences between observed and estimated ARR, the inferential results, comparing the various parameterizations of the prior discounting factor, tend to converge toward the same conclusions in the direction of the generating data effect size starting from a sample size of 150 subjects. All this implies that, for studies conducted on a considerable number of patients, it is possible to tune the prior toward a more informative solution ($d_0 > 0.5$), obtaining results representing a suitable compromise between the available historical information and what is suggested by the data.

In the literature, some reasons are addressed for a suitable discounting of historical prior information. First, the historical data and the current trial evidence may be heterogeneous concerning the study design and conduct [28]. Moreover, as also demonstrated by this simulation analysis, especially for small trials, an informative historical prior may overwhelm the current trial evidence [27].

Another issue outlined in this paper is the potentially misleading information on the treatment effect provided by the posterior median effect for a sample size smaller than 50 patients. This source of bias is evident not only for informative inference but also for low-informative and uninformative analyses. Conversely, the frequentist point estimate is unbiased in terms of the mean because of the proportion estimator's asymptotical unbiasedness over repeated resamples. However, in the frequentist approach, the variability of results across sample replications is very high for small samples, even though the effect, on average, is unbiased [54]. Bayesian estimates, on the other hand, return scenarios of inferential results that are less variable, especially if a minimum amount of historical information is incorporated into the prior.

The frequentist approach considers all the parameters to be fixed; the data are a realization of a random variable. Instead, Bayesian methods assume that all the parameters are random and the data are fixed [54]. This point of view leads to incorporating the available knowledge on the prior parameters into a probability distribution. For this reason, it is important to ensure that the information on which informative priors are based is accurate; otherwise, the resulting estimates and posterior standard deviations could be biased if misleading informative priors are utilized [55].

In this regard, the Bayesian approach leads to thinking about inference in terms of a probability distribution on the treatment effect, rather than a point estimate or confidence interval. Therefore, a Bayesian approach is oriented toward a progressive uncertainty

reduction (on a posterior probability distribution) in treatment effect estimation. Historical information contributes sequentially to the reduction of this uncertainty [56]. The uncertainty can be measured in terms of the CI width. The simulation results demonstrate a narrower CI for small sample sizes (similarly across different true ARRs) for Bayesian analyses compared to the frequentist approach. This effect has also been reported in the literature [57].

The present results show that Bayesian methods can outperform frequentist methods with small samples by providing increased efficiency and an increased ability to determine non-null effects. However, the appropriate prior distribution choice, especially on small datasets, plays a fundamental role. Researchers might need to consult experts, meta-analyses, or review studies in the area of interest to obtain informative, accurate priors that can meaningfully contribute to posterior distributions. Furthermore, a sensitivity analysis on priors (i.e., defining the robustness of conclusions that may be affected by decisions made on the priors) is highly recommended for pediatric trials [14], which is in line with the literature [24] and FDA recommendations [25].

Study Limitations

This study was conducted considering only the conjugate prior beta setting. It may be interesting to explore the impact of inference in the posterior case obtained in a nonclosed form. For example, instead of directly placing a parameter derived on the beta prior, it may be advisable to consider expert elicitation about treatment effects to define the specific prior distribution. Moreover, future research development is needed to investigate the effect of an eventual prior-data conflict on the trial results according to different study size.

5. Conclusions

Bayesian inference is a flexible tool compared to frequentist inference, especially for trials conducted in a poor accrual setting. A full informative Bayesian inference, conducted on small samples, can generate data-insensitive results. On the other hand, the use of an uninformative prior distribution may include, in the final inference, clinically unproven extreme treatment effect hypotheses. A power prior approach on sample sizes smaller than 50 patients seems to be a good compromise between these two methods. However, the choice of parameters and discounting factors should be negotiated with expert pediatricians and should be guided by an appropriate consultation of the scientific literature. In agreement with the FDA recommendations, a sensitivity analysis of priors is highly recommended.

Author Contributions: Conceptualization: D.G. and D.A.; methodology: I.B.; formal analysis: D.A.; writing—original draft preparation: D.A. and D.G.; writing—review and editing: D.A., D.G., G.L., I.B., S.B., and L.D.D.; supervision: D.G. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kadam, R.A.; Borde, S.U.; Madas, S.A.; Salvi, S.S.; Limaye, S.S. Challenges in Recruitment and Retention of Clinical Trial Subjects. *Perspect. Clin. Res.* **2016**, *7*, 137–143. [CrossRef]
2. Pak, T.R.; Rodriguez, M.; Roth, F.P. Why Clinical Trials Are Terminated. *bioRxiv* **2015**, bioRxiv:021543. [CrossRef]
3. Williams, R.J.; Tse, T.; DiPiazza, K.; Zarin, D.A. Terminated Trials in the ClinicalTrials.gov Results Database: Evaluation of Availability of Primary Outcome Data and Reasons for Termination. *PLoS ONE* **2015**, *10*, e0127242. [CrossRef]
4. Rimel, B. Clinical Trial Accrual: Obstacles and Opportunities. *Front. Oncol.* **2016**, *6*, 103. [CrossRef]
5. Mannel, R.S.; Moore, K. Research: An Event or an Environment? *Gynecol. Oncol.* **2014**, *134*, 441–442. [CrossRef]
6. Stensland, K.D.; McBride, R.B.; Latif, A.; Wisnivesky, J.; Hendricks, R.; Roper, N.; Boffetta, P.; Hall, S.J.; Oh, W.K.; Galsky, M.D. Adult Cancer Clinical Trials That Fail to Complete: An Epidemic? *J. Natl. Cancer Inst.* **2014**, *106*, dju229. [CrossRef] [PubMed]
7. Baldi, I.; Lanera, C.; Berchialla, P.; Gregori, D. Early Termination of Cardiovascular Trials as a Consequence of Poor Accrual: Analysis of ClinicalTrials.gov 2006–2015. *BMJ Open* **2017**, *7*, e013482. [CrossRef] [PubMed]
8. Pica, N.; Bourgeois, F. Discontinuation and Nonpublication of Randomized Clinical Trials Conducted in Children. *Pediatrics* **2016**, e20160223. [CrossRef]
9. Baiardi, P.; Giaquinto, C.; Girotto, S.; Manfredi, C.; Ceci, A. Innovative Study Design for Paediatric Clinical Trials. *Eur. J. Clin. Pharmacol.* **2011**, *67*, 109–115. [CrossRef]
10. Greenberg, R.G.; Gamel, B.; Bloom, D.; Bradley, J.; Jafri, H.S.; Hinton, D.; Nambiar, S.; Wheeler, C.; Tiernan, R.; Smith, P.B.; et al. Parents' Perceived Obstacles to Pediatric Clinical Trial Participation: Findings from the Clinical Trials Transformation Initiative. *Contemp. Clin. Trials Commun.* **2018**, *9*, 33–39. [CrossRef]
11. Billingham, L.; Malottki, K.; Steven, N. Small Sample Sizes in Clinical Trials: A Statistician's Perspective. *Clin. Investig.* **2012**, *2*, 655–657. [CrossRef]
12. Kitterman, D.R.; Cheng, S.K.; Dilts, D.M.; Orwoll, E.S. The Prevalence and Economic Impact of Low-Enrolling Clinical Studies at an Academic Medical Center. *Acad. Med. J. Assoc. Am. Med Coll.* **2011**, *86*, 1360–1366. [CrossRef]
13. Joseph, P.D.; Craig, J.C.; Caldwell, P.H. Clinical Trials in Children. *Br. J. Clin. Pharmacol.* **2015**, *79*, 357–369. [CrossRef]
14. Huff, R.A.; Maca, J.D.; Puri, M.; Seltzer, E.W. Enhancing Pediatric Clinical Trial Feasibility through the Use of Bayesian Statistics. *Pediatric Res.* **2017**, *82*, 814. [CrossRef] [PubMed]
15. ICH Topic E 11. Clinical Investigation of Medicinal Products in the Paediatric Population. In *Note for Guidance on Clinical Investigation of Medicinal Products in the Paediatric Population (CPMP/ICH/2711/99)*; European Medicines Agency: London, UK, 2001.
16. European Medicines Agency. *Guideline on the Requirements for Clinical Documentation for Orally Inhaled Products (OIP) Including the Requirements for Demonstration of Therapeutic Equivalence between Two Inhaled Products for Use in the Treatment of Asthma and Chronic Obstructive Pulmonary Disease (COPD) in Adults and for Use in the Treatment of Asthma in Children and Adolescents*; European Medicines Agency: London, UK, 2009.
17. Committee for Medicinal Products for Human Use. *Guideline on the Clinical Development of Medicinal Products for the Treatment of Cystic Fibrosis*; European Medicines Agency: London, UK, 2009.
18. Committee for Medicinal Products for Human Use. *Note for Guidance on Evaluation of Anticancer Medicinal Products in Man*; The European Agency for the Evaluation of Medical Products: London, UK, 1996.
19. O'Hagan, A. Chapter 6: Bayesian Statistics: Principles and Benefits. In *Handbook of Probability: Theory and Applications*; Rudas, T., Ed.; Sage: Thousand Oaks, CA, USA, 2008; pp. 31–45.
20. Azzolina, D.; Berchialla, P.; Gregori, D.; Baldi, I. Prior Elicitation for Use in Clinical Trial Design and Analysis: A Literature Review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1833. [CrossRef]
21. Gajewski, B.J.; Simon, S.D.; Carlson, S.E. Predicting Accrual in Clinical Trials with Bayesian Posterior Predictive Distributions. *Stat. Med.* **2008**, *27*, 2328–2340. [CrossRef] [PubMed]
22. O'Hagan, A. Eliciting Expert Beliefs in Substantial Practical Applications. *J. R. Stat. Soc. Ser. D Stat.* [CD-ROM P55-68]. **1998**, *47*, 21–35.
23. Lilford, R.J.; Thornton, J.; Braunholtz, D. Clinical Trials and Rare Diseases: A Way out of a Conundrum. *BMJ* **1995**, *311*, 1621–1625. [CrossRef]
24. Quintana, M.; Viele, K.; Lewis, R.J. Bayesian Analysis: Using Prior Information to Interpret the Results of Clinical Trials. *Jama* **2017**, *318*, 1605–1606. [CrossRef]
25. Office of the Commissioner; Office of Clinical Policy and Programs. *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials*; CDRH: Rockville, MD, USA; CBER: Silver Spring, MD, USA, 2010. Available online: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials (accessed on 2 February 2021).
26. Gelman, A. Prior Distribution. Chapter 4: Statistical Theory and Methods. In *Encyclopedia of Environmetrics*; Wiley: Hoboken, NJ, USA, 2006.
27. De Santis, F. Power Priors and Their Use in Clinical Trials. *Am. Stat.* **2006**, *60*, 122–129. [CrossRef]
28. Ibrahim, J.G.; Chen, M.-H. Power Prior Distributions for Regression Models. *Stat. Sci.* **2000**, *15*, 46–60.
29. Psioda, M.A.; Ibrahim, J.G. Bayesian Clinical Trial Design Using Historical Data That Inform the Treatment Effect. *Biostat* **2019**, *20*, 400–415. [CrossRef] [PubMed]

Int. J. Environ. Res. Public Health **2021**, *18*, 2095

16 of 16

30.	Ibrahim, J.G.; Chen, M.-H.; Gwon, Y.; Chen, F. The Power Prior: Theory and Applications. *Stat. Med.* **2015**, *34*, 3724–3749. [CrossRef] [PubMed]
31.	Hobbs, B.P.; Carlin, B.P.; Mandrekar, S.J.; Sargent, D.J. Hierarchical Commensurate and Power Prior Models for Adaptive Incorporation of Historical Information in Clinical Trials. *Biometrics* **2011**, *67*, 1047–1056. [CrossRef] [PubMed]
32.	Liu, G.F. A Dynamic Power Prior for Borrowing Historical Data in Noninferiority Trials with Binary Endpoint. *Pharm. Stat.* **2018**, *17*, 61–73. [CrossRef]
33.	Nikolakopoulos, S.; van der Tweel, I.; Roes, K.C.B. Dynamic Borrowing through Empirical Power Priors That Control Type I Error: Dynamic Borrowing with Type I Error Control. *Biom* **2018**, *74*, 874–880. [CrossRef]
34.	Ollier, A.; Morita, S.; Ursino, M.; Zohar, S. An Adaptive Power Prior for Sequential Clinical Trials—Application to Bridging Studies. *Stat. Methods Med Res.* **2020**, *29*, 2282–2294. [CrossRef] [PubMed]
35.	Schmidli, H.; Gsteiger, S.; Roychoudhury, S.; O'Hagan, A.; Spiegelhalter, D.; Neuenschwander, B. Robust Meta-Analytic-Predictive Priors in Clinical Trials with Historical Control Information: Robust Meta-Analytic-Predictive Priors. *Biom* **2014**, *70*, 1023–1032. [CrossRef]
36.	Chuang-Stein, C. An Application of the Beta-Binomial Model to Combine and Monitor Medical Event Rates in Clinical Trials. *Drug Inf. J.* **1993**, *27*, 515–523. [CrossRef]
37.	Zaslavsky, B.G. Bayes Models of Clinical Trials with Dichotomous Outcomes and Sample Size Determination. *Stat. Biopharm. Res.* **2009**, *1*, 149–158. [CrossRef]
38.	Huang, Y.C.; Lin, Y.C.; Wei, C.F.; Deng, W.L.; Huang, H.C. The Pathogenicity Factor HrpF Interacts with HrpA and HrpG to Modulate Type III Secretion System (T3SS) Function and T3ss Expression in *Pseudomonas Syringae* Pv. *Averrhoi. Mol. Plant Pathol.* **2016**, *17*, 1080–1094. [CrossRef]
39.	Shaikh, N.; Shope, T.R.; Hoberman, A.; Muniz, G.B.; Bhatnagar, S.; Nowalk, A.; Hickey, R.W.; Michaels, M.G.; Kearney, D.; Rockette, H.E.; et al. Corticosteroids to Prevent Kidney Scarring in Children with a Febrile Urinary Tract Infection: A Randomized Trial. *Pediatr. Nephrol.* **2020**, *35*, 2113–2120. [CrossRef]
40.	Lehoczky, J.P. Distributions, Statistical: Special and Continuous. In *International Encyclopedia of the Social & Behavioral Sciences*; Elsevier: Amsterdam, The Netherlands, 2001; pp. 3787–3793. ISBN 978-0-08-043076-8.
41.	Wilcox, R.R. A Review of the Beta-Binomial Model and Its Extensions. *J. Educ. Stat.* **1981**, *6*, 3. [CrossRef]
42.	Huang, Y.-Y.; Chen, M.-J.; Chiu, N.-T.; Chou, H.-H.; Lin, K.-Y.; Chiou, Y.-Y. Adjunctive Oral Methylprednisolone in Pediatric Acute Pyelonephritis Alleviates Renal Scarring. *Pediatrics* **2011**, peds-2010. [CrossRef]
43.	De Santis, F. Using Historical Data for Bayesian Sample Size Determination. *J. R. Stat. Soc. Ser. A* **2007**, *170*, 95–113. [CrossRef]
44.	Weber, S. *RBesT: R Bayesian Evidence Synthesis Tools*, Version 1.6-1. Available online: https://cran.r-project.org/web/packages/RBesT/index.html (accessed on 2 February 2020).
45.	Albert, J. *Bayesian Computation with R*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009; ISBN 0-387-92298-9.
46.	Kawasaki, Y.; Shimokawa, A.; Miyaoka, E. Comparison of Three Calculation Methods for a Bayesian Inference of $P\,(\Pi_1 > \Pi_2)$. *J. Mod. Appl. Stat. Methods* **2013**, *12*, 256–268. [CrossRef]
47.	Barry, J. Doing Bayesian Data Analysis: A Tutorial with R and BUGS. *Eur. J. Psychol.* **2011**, *7*, 778–779. [CrossRef]
48.	Lunn, D.; Spiegelhalter, D.; Thomas, A.; Best, N. The BUGS Project: Evolution, Critique and Future Directions. *Stat. Med.* **2009**, *28*, 3049–3067. [CrossRef] [PubMed]
49.	R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.
50.	Geweke, J. *Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments*; Federal Reserve Bank of Minneapolis, Research Department: Minneapolis, MN, USA, 1991; Volume 196.
51.	Bavdekar, S.B. Pediatric Clinical Trials. *Perspect. Clin. Res.* **2013**, *4*, 89–99. [CrossRef] [PubMed]
52.	Gill, D.; Kurz, R. Practical and Ethical Issues in Pediatric Clinical Trials. *Appl. Clin. Trials* **2003**, *12*, 41–45.
53.	Pedroza, C.; Han, W.; Thanh Truong, V.T.; Green, C.; Tyson, J.E. Performance of Informative Priors Skeptical of Large Treatment Effects in Clinical Trials: A Simulation Study. *Stat. Methods Med. Res.* **2018**, *27*, 79–96. [CrossRef] [PubMed]
54.	McNeish, D. On Using Bayesian Methods to Address Small Sample Problems. *Struct. Equ. Modeling Multidiscip. J.* **2016**, *23*, 750–773. [CrossRef]
55.	Depaoli, S. Mixture Class Recovery in GMM under Varying Degrees of Class Separation: Frequentist versus Bayesian Estimation. *Psychol. Methods* **2013**, *18*, 186–219. [CrossRef] [PubMed]
56.	Spiegelhalter, D.J.; Freedman, L.S.; Parmar, M.K.B. Bayesian Approaches to Randomized Trials. *J. R. Stat. Soc. Ser. A* **1994**, *157*, 357. [CrossRef]
57.	Albers, C.J.; Kiers, H.A.L.; van Ravenzwaaij, D. Credible Confidence: A Pragmatic View on the Frequentist vs. Bayesian Debate. *Collabra Psychol.* **2018**, *4*, 31. [CrossRef]