



Article

Sample Size Requirements for Calibrated Approximate Credible Intervals for Proportions in Clinical Trials

Fulvio De Santis [†] and Stefania Gubbiotti ^{*,†}

Dipartimento di Scienze Statistiche, Sapienza University of Rome, Piazzale Aldo Moro n. 5, 00185 Rome, Italy; fulvio.desantis@uniroma1.it

* Correspondence: stefania.gubbiotti@uniroma1.it

† These authors contributed equally to this work.

Abstract: In Bayesian analysis of clinical trials data, credible intervals are widely used for inference on unknown parameters of interest, such as treatment effects or differences in treatments effects. Highest Posterior Density (HPD) sets are often used because they guarantee the shortest length. In most of standard problems, closed-form expressions for exact HPD intervals do not exist, but they are available for intervals based on the normal approximation of the posterior distribution. For small sample sizes, approximate intervals may be not calibrated in terms of posterior probability, but for increasing sample sizes their posterior probability tends to the correct credible level and they become closer and closer to exact sets. The article proposes a predictive analysis to select appropriate sample sizes needed to have approximate intervals calibrated at a pre-specified level. Examples are given for interval estimation of proportions and log-odds.

Keywords: Bayesian inference; highest posterior density intervals; normal approximation; predictive analysis; sample size determination



Citation: De Santis, F.; Gubbiotti, S. Sample Size Requirements for Calibrated Approximate Credible Intervals for Proportions in Clinical Trials. *Int. J. Environ. Res. Public Health* **2021**, *18*, 595. <https://doi.org/10.3390/ijerph18020595>

Received: 1 November 2020

Accepted: 8 January 2021

Published: 12 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The use of Bayesian methods for design, analysis and monitoring of clinical trials is becoming more and more popular. For instance, in some recent contributions [1,2] the Authors note that “compared with its frequentist counterpart, the Bayesian framework has several unique advantages, and its incorporation into clinical trial design is occurring more frequently.” Acknowledgements have been arriving also from official institutions. In 2010 FDA, recognizing the merits of Bayesian inference, authorized and encouraged its use in medical device clinical trials. Similarly Bittle and He observe that “[...] in a major shift, the American College of Cardiology and American Heart Association have recently proposed using Bayesian analysis to create clinical trials guidelines” [3].

There are at least two main motivations for using Bayesian methods. The first is that, unlike frequentist analysis, the Bayesian approach allows the integration of information from a current experiment with pre-trial knowledge. The second advantage is that Bayesian inferential methods are derived from probability distributions that are directly defined on the quantity of interest in the trial (i.e., the parameter). This makes communication between statisticians and experts in the field much more effective than it is when frequentist methods are employed.

With no significant loss of generality, suppose we are concerned with inference on the unknown effect of a new treatment, that we assume to be our parameter of interest. Bayesian methodology is based on elaborations of the posterior distribution of the parameter, which merges pre-experimental knowledge (i.e., the prior distribution) and trial information (i.e., the likelihood function) on this parameter via Bayes theorem. Inferential tools—such as point estimates, set estimates or test statistics—are simply special functionals of the posterior distributions. Nowadays analytic and computational methods for handling complex Bayesian problems are available, even in high dimensional settings.

Nevertheless, the availability of closed-form expressions makes the use of Bayesian analysis more accessible also to non-statisticians. For this reason a relevant part of the available Bayesian literature in clinical trials resorts heavily to normal approximations [4].

Interval estimation is one of the most common techniques used to summarize information on an unknown parameter. Bayesian inference usually relies on exact Highest Posterior Density intervals (HPD). The $(1 - \gamma)$ -HPD interval is the subset of the parameter space of probability $(1 - \gamma)$ whose points have density higher than the density of any value of the parameter outside the interval. When the posterior distribution is symmetric, HPDs are also equal-tails (ET) intervals, i.e., they are limited respectively by the $\gamma/2$ and the $1 - \gamma/2$ quantiles of the posterior density of the parameter. HPDs are, typically, not easy to compute, but of minimal length among intervals of given credibility. For a predictive comparison between HPDs and ETs see [5]. Explicit closed-form expressions for the bounds of common exact credible intervals are in most of the cases, not available even in very common models. However, their computation can be simplified by approximating the exact posterior distribution with a normal density and finding the equal-tails intervals, i.e., the $\gamma/2$ and the $1 - \gamma/2$ quantiles of the approximated (symmetric) normal density.

In many standard models the posterior density has a unique mode internal to its support. The degree of skewness of the posterior distribution with respect to its mode depends on the shapes of the likelihood function and of the prior distribution [6]. As shown in Figure 1 asymmetry affects the quality of approximate credible intervals that in general may differ substantially from exact HPDs. This means that, in general, for approximate intervals: (a) their actual posterior probability is not equal to the nominal credibility of the exact interval; (b) they are not the shortest intervals among those of given posterior probability.

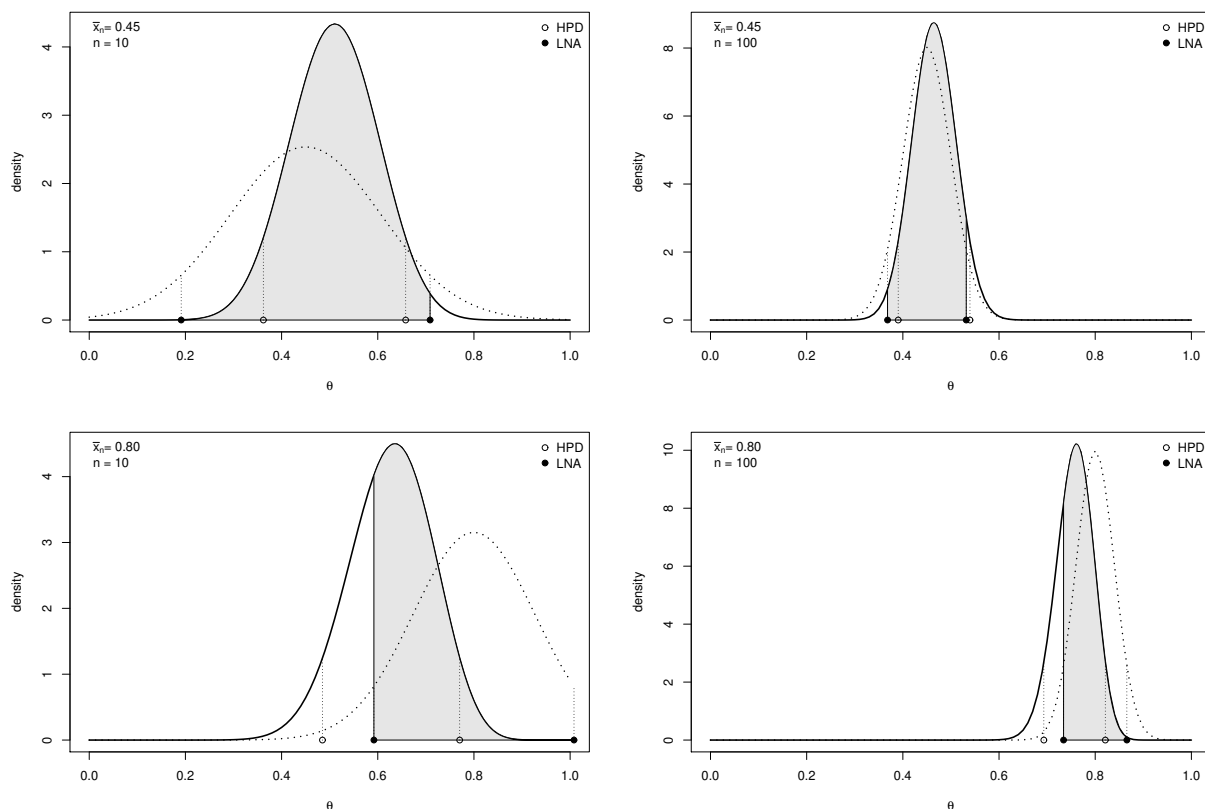


Figure 1. Posterior density, given a prior density of hyperparameters $(\alpha, \beta) = (10.8, 9.2)$, and likelihood approximation, given $\bar{x}_n = 0.45$ (top row) and $\bar{x}_n = 0.8$ (bottom row) for $n = 10$ (left column) and $n = 100$ (right column). Exact credible intervals (HPD: Highest Posterior Density) are denoted by empty circles, likelihood approximated credible intervals (LNA: Likelihood Normal Approximation) are denoted by black circles. The probability that θ belongs to the approximate interval under the exact posterior distribution is highlighted in grey.

Under standard and fairly general conditions [7], the degree of asymmetry of the likelihood function is strictly related to the sample size: as the number of experimental units increases, the shape of the likelihood becomes closer and closer to a Gaussian function whose mode is the maximum likelihood estimate and whose precision is measured by the square root of the observed Fisher Information [8]. Likelihood normalization carries along the same tendency of the posterior distribution and, for sufficiently large sample sizes, the posterior density can be approximated by a normal density with data-dependent parameters. This is the so-called Bayesian Central Limit Theorem. As a consequence, as the sample size increases, exact and approximate intervals become closer and closer and the accuracy of approximate intervals improves.

Example 1 (Single arm phase II trial). Let us consider an example for binary data, where θ is the probability of response to a treatment. The setup, the choice of the prior hyperparameters and a sensitivity analysis will be fully described in Section 4. A Beta prior of mean 0.54 is considered. Figure 1 shows the Beta posterior distributions of θ (solid line) and their normal approximations based on the likelihood (dotted line) for four different data sets. It also reports bounds of approximate intervals (black circle) and of exact HPD intervals (empty circles). Gray areas highlight the probability of the approximate intervals w.r.t. the exact posterior probability distributions. More specifically, when comparing right panels ($n = 100$) and left panels ($n = 10$) better approximations of the posteriors are observed, due to the larger sample size. Furthermore, the comparison between the two rows of panels (sample mean $\bar{x}_n = 0.45$ and $\bar{x}_n = 0.80$, respectively) shows that the distance between the posterior mode and the likelihood mode (i.e., the maximum likelihood estimate) affects the quality of the approximation: in this example, the larger the difference, the greater the discrepancy between exact and approximate intervals.

The problem we discuss in this paper is the selection of the minimal number of observations to obtain approximate sets that are sufficiently accurate. This sample size determination (SSD) problem is addressed from a pre-posterior perspective, i.e., by taking into account the randomness of the posterior density and of credible intervals.

In the existing literature besides a very general introduction to credible intervals [6,7,9,10] one can find reviews on Bayesian SSD in [11–13], articles specifically dedicated to Bayesian SSD using credible intervals in [14–17] and some contributions focused on binomial proportions, such as [18–20]. Recently, methods that take into account the variability of prior opinion have been developed: for instance, some contributions [15,21,22] deal with robustness with respect to the prior distribution, whereas a more recent proposal is about a consensus-based SSD criterion in the presence of a community of priors [23]. The idea of controlling the conflict between alternative procedures is also used for point estimation [24,25].

In the framework of Bayesian SSD based on credible intervals, our innovative purpose is to look for a sample size sufficiently large so that the approximate likelihood interval provides an accurate approximation to the HPD interval determined from the exact posterior distribution of the parameter of interest. It is worth recalling that whereas the HPD interval is obtained from the prior-to-posterior analysis, the likelihood normal approximation is independent on the prior distribution. In this sense our proposed criterion yields the smallest sample size such that the role of the prior in the posterior distribution is made negligible by the information provided by the data. This provides an additional motivation for our proposal, i.e., to find the study dimension that guarantees a substantial equivalence between closed-form formulas based on the normal approximation and exact Bayesian intervals, or, conversely, to evaluate the expected discrepancy between approximate intervals and exact Bayesian intervals.

The paper is organized as follows. In Section 2, after introducing notation, we propose a measure of discrepancy between exact and approximate intervals to be analyzed from a preposterior perspective: we select the minimal sample size so that the expected discrepancy is sufficiently small. Section 3 specifically refers to the Beta-Binomial model when the parameter of interest is the proportion (Section 3.1) and the logodds (Section

3.2) respectively. Section 4 illustrates some numerical examples related to the setup of the phase II clinical trial of Example 1 and makes comparison with other SSD methods. Finally, Section 5 contains some concluding remarks.

2. Methodology

Assume that X_1, X_2, \dots, X_n is a sample from $f_n(\cdot|\theta)$ (either a density or a probability mass function), where $\theta \in \Theta$ is an unknown scalar parameter and Θ is the parameter space. The quantity of interest may be either θ or a relevant function $\psi = g(\theta)$. Following the Bayesian inferential approach, we assume that prior information on θ is available (from experts or from historical data) and converted in a prior probability density function, denoted as $\pi(\cdot)$. Given an observed sample $\mathbf{x}_n = (x_1, x_2, \dots, x_n)$, let

$$\pi(\theta|\mathbf{x}_n) = \frac{f_n(\mathbf{x}_n|\theta)\pi(\theta)}{m(\mathbf{x}_n)}$$

be the posterior distribution of θ , where $m(\mathbf{x}_n) = \int_{\Theta} f_n(\mathbf{x}_n|\theta)\pi(\theta)d\theta$ denotes the marginal distribution of the the data, computed at the observed \mathbf{x}_n . In the following we assume that $\pi(\theta|\mathbf{x}_n)$ has a unique mode.

2.1. Exact and Approximate Intervals

Let $C(\mathbf{x}_n) = [\ell(\mathbf{x}_n), u(\mathbf{x}_n)]$ be an exact credible interval of level $1 - \gamma$, that is a subset of the parameter space such that

$$\mathbb{P}[\theta \in C(\mathbf{x}_n)|\mathbf{x}_n] = 1 - \gamma. \tag{1}$$

In the following, we will focus on HPD intervals. C is HPD if

$$\pi(\theta|\mathbf{x}_n) \geq \pi(\theta'|\mathbf{x}_n), \quad \forall \theta \in C(\mathbf{x}_n) \quad \text{and} \quad \forall \theta' \notin C(\mathbf{x}_n),$$

or, equivalently, if

$$C(\mathbf{x}_n) = \{\theta \in \Theta : \pi(\theta|\mathbf{x}_n) \geq k_\gamma\},$$

where k_γ is such that (1) holds. The values of ℓ and u are the roots of the two equations

$$\pi(\ell|\mathbf{x}_n) = \pi(u|\mathbf{x}_n) \quad \text{and} \quad \int_{\ell}^u \pi(\theta|\mathbf{x}_n)d\theta = 1 - \gamma,$$

and they typically do not have a closed-form expression.

In general, $\pi(\theta|\mathbf{x}_n)$ is not symmetric with respect to its unique mode. Its level of skewness depends on the constitutive elements of Bayesian analysis—the likelihood (i.e., model and observed data) and the prior distribution— and it determines the level of discrepancy between approximate and exact credible intervals. However, as the sample size increases, the shape of both the likelihood function and the posterior density tend to become more and more Gaussian. This happens under standard regularity conditions: (a) the support of the X_i 's does not depend on θ ; (b) the derivatives with respect to θ of likelihood and posterior density at least up to the second order exist; (c) the maximum likelihood estimate of θ , $\hat{\theta}$, is in the interior of the parameter space [6–8]. More specifically, for sufficiently large n we have that

$$\theta|\mathbf{x}_n \approx N[\hat{\theta}, I_n(\hat{\theta})^{-1}], \tag{2}$$

where $I_n(\theta) = -\frac{d^2}{d\theta^2} \ln L(\theta; \mathbf{x}_n)$ is the expected Fisher Information and $L(\theta; \mathbf{x}_n)$ is the likelihood function. Note that this approximation of the posterior distribution does not take into account the prior. From Equation (2) the $(1 - \gamma)$ -likelihood approximate interval for θ is defined as $\tilde{C}(\mathbf{x}_n) = [\tilde{\ell}(\mathbf{x}_n), \tilde{u}(\mathbf{x}_n)]$ where

$$\tilde{\ell} = \hat{\theta} - z_{1-\frac{\gamma}{2}} I_n(\hat{\theta})^{-1/2} \quad \text{and} \quad \tilde{u} = \hat{\theta} + z_{1-\frac{\gamma}{2}} I_n(\hat{\theta})^{-1/2}, \tag{3}$$

with z_ϵ denoting the ϵ -quantile of the standard normal distribution. As a consequence, as n increases, any measure of discrepancy between a chosen feature of exact and approximate intervals tends to become more and more negligible.

When the quantity of interest is $\psi = g(\theta)$, under the same regularity conditions stated above and assuming that the first derivative of g exists and is not equal to 0, the delta method provides the following normal approximation [26]

$$\psi|\mathbf{x}_n \approx N[g(\hat{\theta}), g'(\hat{\theta})^2 I_n(\hat{\theta})^{-1}], \tag{4}$$

and the bounds of the $(1 - \gamma)$ likelihood approximate credible interval for ψ are respectively

$$\tilde{\ell} = g(\hat{\theta}) - z_{1-\frac{\gamma}{2}} \cdot |g'(\hat{\theta})| \cdot I_n(\hat{\theta})^{-1/2} \quad \text{and} \quad \tilde{u} = g(\hat{\theta}) + z_{1-\frac{\gamma}{2}} \cdot |g'(\hat{\theta})| \cdot I_n(\hat{\theta})^{-1/2}. \tag{5}$$

2.2. A Measure of Discrepancy and Predictive Analysis

The set $\tilde{C} = [\tilde{\ell}, \tilde{u}]$ is calibrated if its exact posterior probability is equal to $1 - \gamma$:

$$\mathbb{P}(\theta \in \tilde{C}|\mathbf{x}_n) = F(\tilde{u}|\mathbf{x}_n) - F(\tilde{\ell}|\mathbf{x}_n) = (1 - \gamma), \tag{6}$$

where $F(\cdot|\mathbf{x}_n)$ is the exact posterior cumulative distribution function of the parameter of interest. The departure from this situation can be measured by

$$|\mathbb{P}(\theta \in \tilde{C}|\mathbf{x}_n) - (1 - \gamma)| \tag{7}$$

which quantifies the discrepancy between the actual posterior probability of \tilde{C} (the gray area of each panel of Figure 1 in Example 1) and its nominal value $1 - \gamma$. Notice that, under the typical assumption $0 < \gamma \ll \frac{1}{2}$, this discrepancy takes values in $(0, 1 - \gamma)$. More specifically, it is equal to 0 when \tilde{C} is perfectly calibrated and it is equal to $1 - \gamma$ when $\mathbb{P}(\theta \in \tilde{C}|\mathbf{x}_n) = 0$. Hence, a relative measure based on (7) is

$$P(\mathbf{x}_n) = \frac{|\mathbb{P}(\theta \in \tilde{C}|\mathbf{x}_n) - (1 - \gamma)|}{1 - \gamma} \tag{8}$$

Before observing the data, $P(\mathbf{X}_n)$ is a random object. Therefore the progressive calibration of $\tilde{C}(\mathbf{X}_n)$ can be studied by looking at its expected value

$$e_n^P = \mathbb{E}_d[P(\mathbf{X}_n)],$$

that is computed with respect to the sampling distribution of the data $f_n(\cdot|\theta_d)$ for a design value θ_d . In the following we assume that all the required regularity conditions hold such that the numerical sequence $\{e_n^P, n \in \mathbb{N}\}$ converges to zero.

In order to obtain a calibrated approximate interval, we must select the smallest sample size such that e_n^P is sufficiently small. More formally, for a suitable threshold $\epsilon_P > 0$,

$$n_P^* = \min\{n \in \mathbb{N} : e_n^P < \epsilon_P\}. \tag{9}$$

In some cases the values of e_n^P can be obtained with exact calculations. More often they are obtained via Monte Carlo (MC) simulation. In the latter case, for each sample size n and design value θ_d , we proceed according to the following steps:

- (i) draw N samples $\mathbf{x}_n^{(1)}, \dots, \mathbf{x}_n^{(N)}$ from $f_n(\cdot|\theta_d)$;
 - (ii) compute $\tilde{\ell}(\mathbf{x}_n^{(j)})$ and $\tilde{u}(\mathbf{x}_n^{(j)})$, for $j = 1, \dots, N$;
 - (iii) compute $P(\mathbf{x}_n^{(j)})$, for $j = 1, \dots, N$;
 - (iv) set $e_n^P \simeq \frac{\sum_{j=1}^N P(\mathbf{x}_n^{(j)})}{N}$;
- with a large number of draws, e.g., $N = 10000$.

In the following example, in order to assess the discrepancy between \tilde{C} and C we also consider the absolute distance between their *bounds*

$$B(\mathbf{x}_n) = |\tilde{\ell}(\mathbf{x}_n) - \ell(\mathbf{x}_n)| + |\tilde{u}(\mathbf{x}_n) - u(\mathbf{x}_n)|$$

and we compare n_p^* with

$$n_B^* = \min\{n \in \mathbb{N} : e_n^B < \epsilon_B\}, \quad (10)$$

where

$$e_n^B = \mathbb{E}_d[B(\mathbf{X}_n)],$$

and $\epsilon_B > 0$ is a chosen threshold. Note that, unlike $P(\mathbf{x}_n)$ (and e_n^P), the discrepancy $B(\mathbf{x}_n)$ (and e_n^B) depends on the unit of measurement of the data and its range is case-specific. Therefore the choice of ϵ_B is a critical issue, unless the parameter space is bounded (as in Example 1 where the parameter space is $(0, 1)$). Similar measures of discrepancy based on the bounds of credible intervals have been recently proposed [23].

3. Examples: The Beta-Binomial Model

In order to illustrate the ideas sketched above we now consider an example within the Beta-Binomial model. Let $X_i|\theta \sim \text{Ber}(\theta)$, $i = 1, \dots, n$ (i.i.d.), $\theta \in (0, 1)$ and $\theta \sim \text{Be}(\alpha, \beta)$, $\alpha, \beta > 0$. Then, from standard results [6], $\theta|\mathbf{x}_n \sim \text{Be}(\tilde{\alpha}, \tilde{\beta})$, where $\tilde{\alpha} = \alpha + s_n$, $\tilde{\beta} = \beta + n - s_n$ and $s_n = \sum_{i=1}^n x_i$. In the following we first analyze credible intervals for θ and then for the log-odds $\psi = g(\theta) = \ln \frac{\theta}{1-\theta}$.

3.1. Credible Intervals for a Proportion

In this model exact HPD credible intervals for θ do not have closed-form expressions. However, HPD bounds are easily obtained using the `hdi()` function of the `HDInterval` package of R, [27], which simply requires the R function `qbeta()` in input. Conversely, closed-form expressions for approximate intervals are easily obtained as follows. Recalling that $\hat{\theta} = \bar{x}_n$ and $I_n(\theta) = \frac{n}{\theta(1-\theta)}$, from Equation (3) the bounds of the *likelihood approximate interval* are

$$\tilde{\ell} = \bar{x}_n - z_{1-\frac{\gamma}{2}} \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} \quad \text{and} \quad \tilde{u} = \bar{x}_n + z_{1-\frac{\gamma}{2}} \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}.$$

3.2. Credible Intervals for the Log-Odds

As before, exact credible intervals for ψ do not have a closed-form expression. HPD bounds can be obtained via MC simulation as follows:

- (i) draw $\theta^{(1)}, \dots, \theta^{(M)}$ from the posterior Beta density, where M is a large number;
- (ii) compute $\psi^{(j)} = g(\theta^{(j)})$, for $j = 1, \dots, M$;
- (iii) use the R function `HDInterval::hdi` with the MC draws $\psi^{(1)}, \dots, \psi^{(M)}$ in input.

Closed-form expression of approximate credible intervals for ψ are obtained from Equation (5) noting that

$$g(\hat{\theta}) = \ln \frac{\bar{x}_n}{1-\bar{x}_n} \quad \text{and} \quad g'(\hat{\theta}) = \frac{1}{\bar{x}_n(1-\bar{x}_n)}.$$

Specifically, we have

$$\tilde{\ell} = \ln \frac{\bar{x}_n}{1-\bar{x}_n} - z_{1-\frac{\gamma}{2}} \cdot \sqrt{\frac{1}{n\bar{x}_n(1-\bar{x}_n)}} \quad \text{and} \quad \tilde{u} = \ln \frac{\bar{x}_n}{1-\bar{x}_n} + z_{1-\frac{\gamma}{2}} \cdot \sqrt{\frac{1}{n\bar{x}_n(1-\bar{x}_n)}}.$$

Note that in the Beta-Binomial model the values of e_n^P can be obtained using either exact calculations or MC simulations as described in Section 2.2.

4. Application to Clinical Trials

Let us assume that in an early phase trial we are interested in estimating the rate of response, θ , to an experimental treatment using a credible interval. As in Example 1 we consider the setup of a single-arm phase II trial. Specifically, the goal of the study is to test the combination of lenalidomide and rituximab in patients with recurrent indolent non-follicular lymphoma [28–30]. The endpoint is the overall response rate $\hat{\theta}$, that is the proportion of eligible patients who achieved complete, unconfirmed or partial response.

In the trial conducted between 2009 and 2011, 21 responses were observed out of 39 eligible patients. These historical data are used to elicit a Beta prior density for θ . More specifically, we set the prior mean equal to $\alpha/(\alpha + \beta) = 0.54$ and we consider several values for the prior sample size (i.e., the amount of information contained in the prior) that for the Beta model is $\alpha + \beta$ [31]. For illustrative purposes in the following example we set $\alpha + \beta$ equal to 5, 10 and 20. Moreover, for comparison, we also consider a uniform density as non-informative prior (e.g., $\alpha = \beta = 1$). The design value θ_d is set equal to 0.45, that is the lowest acceptable value for the overall response rate [28]. In order to evaluate the impact of the design parameter we also consider $\theta_d = 0.8$ that represents a much more optimistic design scenario.

Figure 2 shows the behaviour of e_n^P for increasing values of the sample size n under different prior assumptions. Table 1 reports the optimal sample sizes n_p^* and n_B^* obtained using criteria (9) and (10) for several choices of the prior hyperparameters, when $\theta_d = 0.45$ and $\theta_d = 0.8$, given $\epsilon_P = \epsilon_B = 0.01$ (i.e., 1% of the width of the parameter space). Table 1 also contains the optimal sample sizes obtained using the Average Length Criterion ALC [13], given a threshold for the interval width as small as 0.1, for both exact (n_L^*) and approximate intervals (n_L^*).

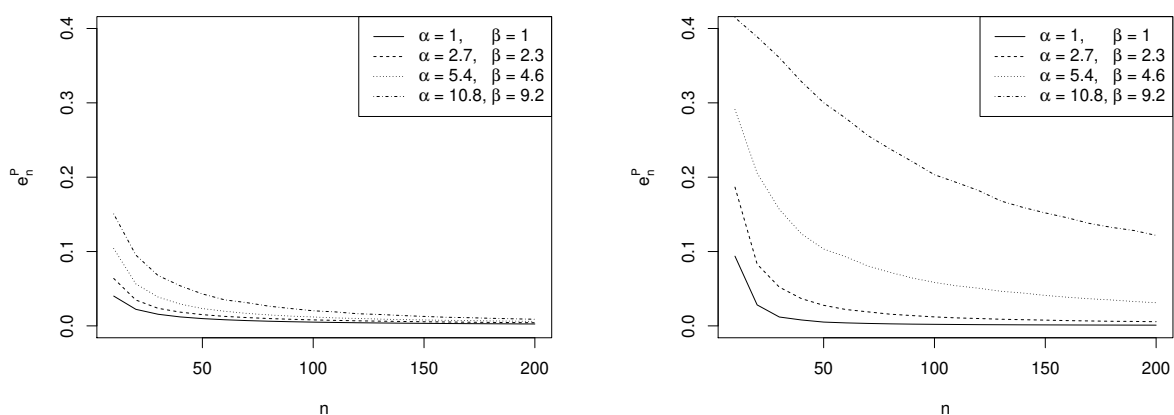


Figure 2. Plots of e_n^P as a function of n for several values of the prior hyperparameters (α, β) , with $\theta_d = 0.45$ (left column) and $\theta_d = 0.8$ (right column).

The most relevant comments are the following.

1. *Effect of sample size.* As expected, the values of e_n^P decrease as n increases and depend on the specific choices of α , β and θ_d as commented in the following remarks.
2. *Effect of prior sample size.* For each value of n , the larger $\alpha + \beta$, the greater the values of e_n^P . In fact, as the prior becomes more and more concentrated around the prior mean 0.54, the weight of the prior in the posterior distribution increases with respect to the role of the likelihood. This makes the discrepancy between Bayesian exact intervals and their likelihood approximation more striking. Moreover, when the uniform non-informative prior is considered, the smallest values of e_n^P are observed (see solid line in Figure 2). As a consequence, larger values of the prior sample size imply greater values of n_p^* , as shown in Table 1.
3. *Effect of the difference between design value and prior mean.* When the distance between θ_d and the prior mean $\alpha/(\alpha + \beta)$ is relatively large and, at the same time, the prior

sample size $\alpha + \beta$ dominates n , the posterior mode and the maximum likelihood estimate are well separated. In other words, Equation (4) does not provide a good approximation of the posterior density of θ . This explains the larger values of e_n^P , in the right panel of Figure 2, where $|\theta_d - \mathbb{E}(\theta)| = 0.35$, with respect to those observed in the left panel, where $|\theta_d - \mathbb{E}(\theta)| = 0.09$. As before, the effect of the difference between design value and prior mean on e_n^P also reflects on the values of the optimal sample sizes reported in Table 1. For instance, under the most informative prior, if $|\theta_d - \mathbb{E}(\theta)| = 0.09$, then $n_p^* = 182$; conversely, when $|\theta_d - \mathbb{E}(\theta)| = 0.35$, a huge number of experimental units (e.g., $n_p^* = 2911$) is required to have a sufficiently small expected discrepancy.

4. *Comparison with n_B^* .* As expected, the trend of n_B^* w.r.t. to (α, β) and θ_d is consistent with that of n_p^* .
5. *Comparison with ALC.* For each θ_d , n_L^* becomes slightly smaller when the prior sample size gets larger and the corresponding posterior is more concentrated (see Table 1). Conversely, since approximate intervals do not depend on the prior, n_L^* is not affected by the choice of prior hyperparameters. Furthermore, when the design value is closer to the boundary of the parameter space, the posterior distribution and, consequently, its approximation, become more concentrated, yielding shorter intervals. Hence the values of n_L^* and of n_L^* are uniformly smaller for $\theta_d = 0.80$ than for $\theta_d = 0.45$. It is interesting to note the opposite impact of the prior sample size $\alpha + \beta$ on n_p^* and n_B^* on the one hand, and on n_L^* on the other hand. In fact, larger values of $\alpha + \beta$ determine shorter intervals and smaller values of n_L^* . On the contrary, when $\theta_d \neq \mathbb{E}(\theta)$, a more concentrated prior implies a more remarkable discrepancy between the posterior and its likelihood approximation and, consequently, yields greater values of n_p^* and n_B^* .

Table 1. Optimal sample sizes for several choices of the prior hyperparameters and of the design values, given $\epsilon_P = \epsilon_B = 0.01$ and $\epsilon_L = 0.1$.

θ_d	(α, β)	(1, 1)	(2.7, 2.3)	(5.4, 4.6)	(10.8, 9.2)
0.45	n_p^*	49	80	119	182
	n_B^*	42	96	180	347
	n_L^*	265	262	257	247
	n_L^*	267	267	267	267
0.80	n_p^*	35	118	646	2911
	n_B^*	91	228	482	992
	n_L^*	170	169	169	167
	n_L^*	172	172	172	172

One of the drawbacks of approximate intervals for θ is that it is not guaranteed that $(\tilde{\ell}, \tilde{u}) \subseteq [0, 1]$. A common solution in the applications is to transform the parameter into the log odds scale so that the normal approximation of the posterior improves. As an example we implemented the credible intervals introduced in Section 3.2. Figure 3 shows the behavior of e_n^P as a function of n for the same choices of hyperparameters and design values used in the previous example. Similar remarks apply.

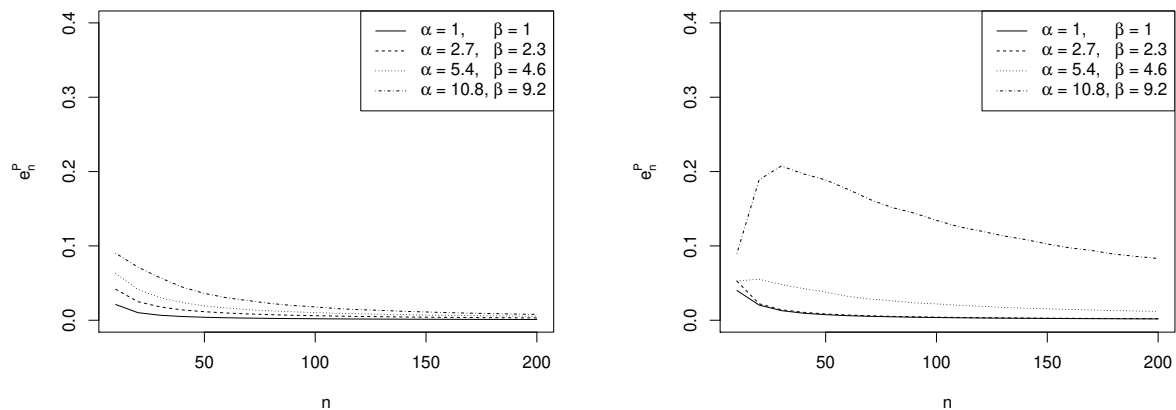


Figure 3. Plots of e_n^P as a function of n for several values of the prior hyperparameters (α, β) with $\theta_d = 0.45$ (left panel) and $\theta_d = 0.8$ (right panel), when the logodds ψ is the parameter of interest.

5. Conclusions

The control of relevant aspects of interval estimates is the starting point for the definition of several SSD criteria both from the frequentist and from the Bayesian perspective. For instance, in the Bayesian side, traditional criteria rely on the pre-posterior control of length and position of credible intervals. In this article we focus on a different request: we look for a sample size sufficiently large so that the approximate likelihood interval provides an accurate approximation to the HPD interval determined from the exact posterior distribution of the parameter of interest. Since the likelihood normal approximation does not depend on the prior distribution, another way to interpret the criterion is that it provides the smallest sample size such that the role of the prior in the posterior distribution is made negligible by the information provided by the data. This kind of analysis can be read in two different ways. On the one side, one can know the number of units needed to use safely closed-form and handy formulas (those provided by the normal approximation) in the place of exact Bayesian intervals. On the other hand, a data analyst who uses approximate intervals instead of exact Bayesian intervals can know the price of this choice in terms of expected discrepancy.

From another perspective this kind of preposterior analysis allows one to know what the study dimension should be for a consensus between a Bayesian interval and a frequentist interval, i.e., a non-informative analysis.

In general, the criterion we propose does not control the main goal of a clinical trial, that can be, for instance, accuracy of estimation or efficacy/inefficacy of a given treatment. For this reason, our criterion should be put beside additional criteria specifically related to the main goal of the trial. For instance in our examples of Section 4 we consider the optimal sample sizes based on ALC. Then, taking the maximum between the two sample sizes obtained using the two criteria, one can control both interval length and accuracy of approximation.

Possible extensions of this work are listed below.

1. Other models. The methodology proposed in the paper can be easily extended to other models and setups relevant to clinical trials applications. A natural extension is to two-arms designs for the comparison of two proportions (difference or log odds ratio), in which the additional issue of units allocation arises [32]. For a predictive approach to allocation based on the control of posterior variances, see for instance [33]. See also [5] for related ideas in the Poisson model.
2. Probability vs. Expectation. In Section 2.2 we propose to summarize the predictive distribution of the discrepancy using the expected value w.r.t. $f_n(\cdot|\theta_d)$. An alternative is to take into account the whole probability distribution of P and to determine the smallest n such that $\mathbb{P}[P(\mathbf{X}_n) > \epsilon_P]$ is sufficiently small.

3. Design prior. For simplicity in this article we have performed preposterior calculations using the sampling distribution $f_n(\cdot|\theta_d)$. An alternative is to consider the so-called *two-priors approach* [23,24,30,34]) which avoids local optimality by replacing the design value with the design prior.
4. Decision-theoretic approach. The approach proposed in the paper is performance-based. Alternatively one could follow some previous works and rephrase the problem in a decision-theoretic framework and define a measure of discrepancy based on the posterior expected loss of C and \tilde{C} . We will elaborate on this in the future.

Author Contributions: Conceptualization, F.D.S.; Data curation, S.G.; Formal analysis, F.D.S. and S.G.; Methodology, F.D.S. and S.G.; Software, S.G.; Visualization, S.G.; Writing—original draft, F.D.S. and S.G.; Writing—review and editing, F.D.S. and S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors would like to thank the guest editors of this Special Issue and the reviewers.

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

References

1. Lee, J.; Chu, C.T. Bayesian clinical trials in action. *Stat. Med.* **2012**, *31*, 2955–2972.
2. Yin, G.; Lam, C.K.; Shi, H. Bayesian randomized clinical trials: From fixed to adaptive design. *Contemp. Clin. Trials* **2017**, *59*, 77–86, doi:10.1016/j.cct.2017.04.010.
3. Bittl, J.A.; He, Y. Bayesian analysis. A practical approach to interpret clinical trials and create clinical practice guidelines. *Circ. Cardiovasc. Qual. Outcomes* **2017**, *10*, e003563.
4. Spiegelhalter, D.J.; Abrams, K.R.; Myles, J.P. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*; Statistics in Practice; Wiley: Chichester, UK, 2004.
5. De Santis, F.; Gubbiotti, S. A note on the progressive overlap of two alternative Bayesian intervals. *Commun. Stat. Theory Methods* **2019**, 1–18, doi:10.1080/03610926.2019.1692034.
6. Lesaffre, E.; Lawson, A.B. *Bayesian Biostatistics*; Wiley: Chichester, UK, 2012.
7. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehtari, A.; Rubin, D.B. *Bayesian Data Analysis*, 3rd ed.; Chapman & Hall/CRC Texts in Statistical Science; Taylor & Francis: Boca Raton, FL, USA, 2013.
8. Kalbfleisch, J.G. *Probability and Statistical Inference. Volume 2: Statistical Inference*, 2nd ed.; Springer: New York, NY, USA, 1985.
9. Robert, C.P. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed.; Springer: New York, NY, USA, 2007.
10. Meeker, W.Q.; Hahn, G.J.; Escobar, L.A. *Statistical Intervals: A Guide for Practitioners and Researchers*; Wiley: Hoboken, NJ, USA, 2017.
11. Brutti, P.; De Santis, F.; Gubbiotti, S. Bayesian—Frequentist sample size determination: A game of two priors. *Metron* **2014**, *72*, 133–151.
12. Adcock, C.J. Sample size determination: A review. *J. R. Stat. Soc. Ser. D Stat.* **1997**, *46*, 261–283, doi:10.1111/1467-9884.00082.
13. Joseph, L.; du Berger, R.; Belisle, P. Bayesian and mixed Bayesian/likelihood criteria for sample size determination. *Stat. Med.* **1997**, *16*, 769–781.
14. Joseph, L.; Wolfson, D. Interval-based versus decision theoretic criteria for the choice of sample size. *J. R. Stat. Soc. Ser. Stat.* **1997**, *46*, 145–149.
15. Brutti, P.; De Santis, F. Robust Bayesian sample size determination for avoiding the range of equivalence in clinical trials. *J. Stat. Plan. Inference* **2008**, *138*, 1577–1591, doi:10.1016/j.jspi.2007.05.041.
16. Cao, J.; Lee, J.J.; Alber, S. Comparison of Bayesian sample size criteria: ACC, ALC, and WOC. *J. Stat. Plan. Inference* **2009**, *139*, 4111–4122, doi:10.1016/j.jspi.2009.05.041.
17. Gubbiotti, S.; De Santis, F. A Bayesian method for the choice of the sample size in equivalence trials. *Aust. New Zealand J. Stat.* **2011**, *53*, 443–460.
18. Joseph, L.; Wolfson, D.B.; Berger, R.D. Sample Size Calculations for Binomial Proportions via Highest Posterior Density Intervals. *J. R. Stat. Soc. Ser. D Stat.* **1995**, *44*, 143–154.

19. M'Lan, C.E.; Joseph, L.; Wolfson, D.B. Bayesian sample size determination for binomial proportions. *Bayesian Anal.* **2008**, *3*, 269–296, doi:10.1214/08-BA310.
20. De Santis, F.; Fasciolo, M.C.; Gubbiotti, S. Predictive control of posterior robustness for sample size choice in a Bernoulli model. *Stat. Methods Appl.* **2013**, *22*, 319–340, doi:10.1007/s10260-012-0225-0.
21. De Santis, F. Sample size determination for robust Bayesian analysis. *J. Am. Stat. Assoc.* **2006**, *101*, 278–291.
22. Brutti, P.; De Santis, F.; Gubbiotti, S. Robust Bayesian sample size determination in clinical trials. *Stat. Med.* **2008**, *27*, 2290–2306, doi:10.1002/sim.3175.
23. Joseph, L.; Belisle, P. Bayesian consensus-based sample size criteria for binomial proportions. *Stat. Med.* **2019**, doi:10.1002/sim.8316.
24. Brutti, P.; De Santis, F.; Gubbiotti, S. Predictive measures of the conflict between frequentist and Bayesian estimators. *J. Stat. Plan. Inference* **2014**, *148*, 111–122, doi:10.1016/j.jspi.2013.12.009.
25. De Santis, F.; Gubbiotti, S. A decision-theoretic approach to sample size determination under several priors. *Appl. Stoch. Model. Bus. Ind.* **2017**, *33*, 282–295.
26. Casella, G.; Berger, R. *Statistical Inference*; Duxbury: Belmont, CA, USA, 2001.
27. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
28. Sacchi, S.; Marcheselli, R.; Bari, A.; Buda, G.; Molinari, A.L.; Baldini, L.; Vallisa, D.; Cesaretti, M.; Musto, P.; Ronconi, S.; et al. Safety and efficacy of lenalidomide in combination with rituximab in recurrent indolent non-follicular lymphoma: final results of a phase II study conducted by the Fondazione Italiana Linfomi. *Haematologica* **2016**, *101*, e196–e199.
29. Zhou, H.; Lee, J.J.; Yuan, Y. BOP2: Bayesian optimal design for phase II clinical trials with simple and complex endpoints. *Stat. Med.* **2017**, *36*, 3302–3314.
30. Sambucini, V. Bayesian predictive monitoring with bivariate binary outcomes in phase II clinical trials. *Comput. Stat. Data Anal.* **2019**, *132*, 18–30.
31. Morita, S.; Thall, P.F.; Muller, P. Determining the effective sample size of a parametric prior. *Biometrics* **2008**, *64*, 595–602, doi:10.1111/j.1541-0420.2007.00888.x.
32. M'Lan, C.E.; Joseph, L.; Wolfson, D.B. Bayesian Sample Size Determination for Case-Control Studies. *J. Am. Stat. Assoc.* **2006**, *101*, 760–772, doi:10.1198/016214505000001023.
33. De Santis, F.; Perone Pacifico, M.; Sambucini, V. Optimal predictive sample size for case-control studies. *Appl. Stat.* **2004**, *53*, 427–441.
34. Wang, F.; Gelfand, A.E. A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statist. Sci.* **2002**, *17*, 193–208, doi:10.1214/ss/1030550861.