Quantification of prior impact in terms of effective current sample size

Manuel Wiesenfarth | Silvia Calderazzo

Division of Biostatistics, German Cancer Research Center, Heidelberg, Germany

Correspondence
Manuel Wiesenfarth, Division of Biostatistics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. Email: m.wiesenfarth@dkfz.de

Abstract
Bayesian methods allow borrowing of historical information through prior distributions. The concept of prior effective sample size (prior ESS) facilitates quantification and communication of such prior information by equating it to a sample size. Prior information can arise from historical observations; thus, the traditional approach identifies the ESS with such a historical sample size. However, this measure is independent of newly observed data, and thus would not capture an actual “loss of information” induced by the prior in case of prior-data conflict. We build on a recent work to relate prior impact to the number of (virtual) samples from the current data model and introduce the effective current sample size (ECSS) of a prior, tailored to the application in Bayesian clinical trial designs. Special emphasis is put on robust mixture, power, and commensurate priors. We apply the approach to an adaptive design in which the number of recruited patients is adjusted depending on the effective sample size at an interim analysis. We argue that the ECSS is the appropriate measure in this case, as the aim is to save current (as opposed to historical) patients from recruitment. Furthermore, the ECSS can help overcome lack of consensus in the ESS assessment of mixture priors and can, more broadly, provide further insights into the impact of priors. An R package accompanies the paper.

KEYWORDS
Bayesian adaptive clinical trial design, prior-data conflict, prior effective sample size, prior elicitation, prior information, robust priors

1 | INTRODUCTION

The elicitation of a prior distribution in a Bayesian data analysis represents both an opportunity to incorporate relevant prior information and a challenge, as the amount and impact of such information may be difficult to fully evaluate a priori. Two key concepts in this respect are those of prior informativeness and prior-data conflict. Quantification of such characteristics in an easily interpretable form is of major importance, as it facilitates communication and transparency of the prior choice. For these reasons, the concept of the prior effective sample size (prior ESS), which equates prior information to a certain number of observations, has gained increasing popularity in the context of clinical trial design.

The possibility of incorporating historical information in clinical trials is of particular relevance, for example, in the context of rare diseases, extrapolation of adult information in pediatric trials (see, eg, Gamalo-Siebers et al., 2017), or when treatments need to be evaluated in a narrow time frame such...
as in case of the 2013 to 2016 Ebola outbreak (Brueckner et al., 2018). As the possible number of samples is usually restricted, quantification of the amount of information induced by the prior is important to avoid domination of the prior information on posterior inference. Quantification of prior informativeness in terms of the prior ESS is then particularly useful: the prior is viewed as arising from a historical trial in which a certain number of observations were collected. Bayes’ rule allows incorporating such observations sequentially in the analysis, by considering the posterior distribution obtained in the historical trial as the prior distribution for the current trial. Well-known analytical results are available for conjugate priors of exponential family likelihoods. Morita et al. (2008) proposed an algorithmic extension applicable to arbitrary prior-likelihood settings. In the following, we assume prior information to arise from historical data, but of course it can correspond to any external information. The criteria for the selection of historical trials for prior elicitation were proposed in Pocock (1976). However, if, despite compliance with carefully elicited criteria, the information collected in the historical trial is not consistent (commensurate) with the information collected during the current trial, a prior-data conflict is observed, and the consequences on the final analysis may be severe.

The assessment of prior-data conflict was investigated in the context of model checking, jointly with the assessment of the data parametric model, in several works (see, eg, Box, 1980; Gelman et al., 1996). More recently, authors have focused on methodologies that allow to check for prior-data conflict independently of the data distributional assumptions (see, eg, Evans and Moshonov, 2006; Bousquet, 2008). The prior ESS can quantify the strength of prior information, but it is not designed to detect and/or quantify prior-data conflict. In a normal prior-normal likelihood model, for example, all priors sharing the same variance have the same prior ESS, but their impact in terms of posterior bias (and thus mean squared error [MSE]) can vary substantially depending on their location with respect to the observed data. The need for a proper understanding of the prior in light of the current data likelihood is advocated, for example, in Gelman et al. (2017). The work of Reimherr et al. (2014) points in this direction in that it computes the effective sample size in terms of samples from the current data model (ie, in a clinical trial context, samples with characteristics consistent with the current trial). Under extreme prior-data conflict, the prior may account for a negative number of samples, showing that information is subtracted, rather than added, by the elicited prior. Based on their ideas, we construct the novel measure prior effective current sample size (ECSS), tailored to the context of clinical trials but also of possible broader interest. Roughly, it quantifies the number of current samples to be added or subtracted to the likelihood in order to obtain a posterior inference equivalent to that of a baseline prior model (eg, in terms of MSE).

The ECSS measure may be of particular relevance for quantification of the impact of priors which dynamically borrow prior information, that is, adaptively discard prior information in case of prior-data conflict. Popular choices of such priors are power priors (eg, Ibrahim et al., 2015; Gravestock and Held, 2017), commensurate priors (eg, Hobbs et al., 2012) and (robust) mixture priors (eg, Berger and Berliner, 1986; Schmidli et al., 2014). Different approaches have been proposed to estimate the ESS of mixture priors, but the difficulty of this task suggests that “an entirely convincing approach to ESS is missing” (Neuenschwander and Schmidli, 2019). We argue that the proposed ECSS may be more suitable when a data-dependent measure is sought.

Further, the ECSS can be of particular interest in adaptive trials where the target number of randomized control patients is reduced by the number of individuals accounted for by the prior, as estimated at an interim analysis. Examples of such designs have been studied by Hobbs et al. (2013), Schmidli et al. (2014), Bennett (2018), and Chen et al. (2018). As the interest effectively lies in savings in terms of current, rather than historical, randomized patients, this application seems very suitable for the proposed ECSS.

The paper is organized as follows. Section 2.1 provides an overview of the prior ESS, while Section 2.2 details the proposed ECSS. The two measures are then applied to robust mixture, commensurate and power priors and contrasted in Section 3. The practical usefulness of the ECSS is further demonstrated in a simulation study of an adaptive trial design in Section 4. In the following, we focus on the one-parameter case for ease of notation, although all approaches can also handle multiple parameters.

2 | PRIOR ESS: TWO FRAMEWORKS

2.1 | Prior ESS

The traditional data-independent approach to ESS is well defined and widely known for models arising from exponential family likelihoods with conjugate priors. Closed-form results follow from the fact that conjugacy allows to view a prior as a posterior from a historical analysis, performed in turn under a baseline (vague) prior belonging to the same distribution. The prior ESS with respect to historical samples (prior ESS) can thus be easily derived as a function of the historical posterior/current prior parameters. A well-known example is that of a Beta($s_1, s_2$) prior in a beta-binomial model. Such
prior can arise as a posterior of an analysis in which $s_1$ successes and $s_2$ failures have been observed, under a baseline prior $\text{Beta}(\alpha, \beta)$, with $\alpha$ and $\beta$ arbitrarily small. This gives rise to the notion of $s_1 + s_2$ as the prior ESS. Another example concerns the normal prior with variance $\sigma^2$ for the mean of independent and identically distributed (i.i.d.) normal data with variance $\sigma^2$, where the prior ESS can be deduced to equal $\sigma^2 / \sigma^2$. The approach of Morita et al. (2008) provides a generalization of this reasoning. Let $\pi$ be the prior of interest with mean $\theta_\pi$ and variance $\sigma^2_\pi$, a vector of $n$ observations having probability density function $f_n(y_{1:n} | \theta)$ indexed by the parameter $\theta$, and let $f_n(y_{1:n} | \pi) = \int f_n(y_{1:n} | \theta) \pi(\theta) d\theta$ be the marginal data distribution. Define an $\varepsilon$-information prior (a prior with prior ESS approximately equal to zero), $\pi_0$, as a prior having the same parametric family and mean as $\pi$, but arbitrarily large variance (subjected to the parametric family constraint) $\sigma^2_0$. Moreover, let $D_{\text{curv}}(\pi(\theta))$ denote the curvature of the log of density $\pi$, $D_{\text{curv}}(\pi(\theta)) = -d^2 \log(\pi(\theta))/d\theta^2$, which quantifies the information contained in the prior $\pi$, and analogously $D_{\text{curv}}(\pi(\theta) | y_{1:n})$ for the posterior distribution induced by the prior $\pi_0$ and the data $y_{1:n}$. The prior ESS is then given by sample size $m$ which minimizes

\[
\text{prior ESS} = \arg \min_m \left[ \int D_{\text{curv}}(\pi_0(\theta) | y_{1:m}) f_m(y_{1:m}) dy_{1:m} \right] - D_{\text{curv}}(\pi(\theta)) \bigg| \theta = \theta_\pi.
\]

A schematic representation of the algorithm is provided on the left panel of Figure 1.

Note that the prior distribution which eventually generates the observed data is the same as the prior of interest, that is, no prior-data conflict is assumed. Alternatives to the curvature measure can be considered, although Morita et al. (2008) finds that such measure allows to match known analytical results. We speculate that the reason for this is that the curvature of the log-density has a close resemblance to the Fisher information obtained from the likelihood, which is proportional to the sample size for i.i.d. observations. However, as for the Fisher information itself, there are cases in which it may not be a sensible measure of informativeness, for example, for mixture distributions, as discussed in Section 3.1.

2.2 Effective current sample size

The prior ESS provides a formal and intuitive quantification of the amount of information introduced by the chosen prior and serves as a valuable data-independent tool for prior elicitation. However, it might not give the full picture of the prior impact in case of prior-data conflict. A data-dependent algorithm in the same spirit as Morita et al. (2008), but accounting for potential discrepancy between the data and the prior of interest, has been proposed by Reimherr et al. (2014). Reimherr et al.’s (2014) algorithm compares two posteriors, rather than a prior and a posterior, and replaces the curvature as a measure of information by a measure which also captures prior-data conflict. We build on their work to propose the ECSS of a prior, which is tailored to the application in the context of clinical trial design.

Specifically, let $\pi$ be again the prior of interest with mean $\theta_\pi$, $\pi_0$ a baseline prior (an objective or reference prior; Kass and Wasserman, 1996; Berger et al., 2009) and $f_n(y_{1:n} | \theta_0)$ be the data distribution. To adopt a measure commonly of interest in clinical trial and estimation settings, we target the MSE measure induced by the posterior mean estimate $E_{\theta}|(\theta | y_{1:k})$, that is, $D_{\text{MSE}}(\theta) = E_{\theta_0}|(\theta_0 | y_{1:k}) - \theta_0^2$. The true parameter value $\theta_0$ is either known in case of a prospective use at planning stage, or replaced by its estimator $\hat{\theta}_0 = E_{m_0}(\theta | y_{1:n})$, that is, the posterior mean

![FIGURE 1](https://example.com/figure1.png)

**FIGURE 1** Schematic representation of the algorithm by Morita et al. (2008) (left) and the algorithm for the computation of the ECSS (right). ECSS, effective current sample size; MSE, mean square error [This figure appears in color in the electronic version of this article, and any mention of color refers to that version]
under the baseline prior specification given all observed data, in case of use in conjunction with observed data. We then define the ECSS at target sample size $k$ as the sample size $m$ which minimizes

$$ECSS = \arg \min_{m} D_{\text{MSE}}^{\theta_0} \left[ \pi(\theta | y_{1:k}) \right] - D_{\text{MSE}}^{\theta_0} \left[ \pi_\theta(\theta | y_{1:k}) \right].$$

That is, we aim at quantifying, at a given sample size and for a given target measure, the number of samples to be added or subtracted to the informative prior model in order to reach the same inferences as the baseline prior model, assuming that the information at hand (the posterior mean under the baseline prior) provides our best guess of the truth. The ECSS is in line with Reimherr et al.’s (2014) notion that the information contained in a posterior is the sum of the information contained in the prior and in the likelihood, and thus an informative prior can be easily equated to a certain likelihood sample size given a posterior under a baseline prior (which only contains information coming from the likelihood).

The right panel of Figure 1 provides a schematic representation of the algorithm and Figure 2 illustrates an example where normal data $y \sim N(1, 2^2)$ and a baseline prior $\mu \sim N(0, 10^2)$ are assumed. The nonconflicting prior $\mu \sim N(1.3, 0.5^2)$ and the conflicting prior $N(2.5, 0.5^2)$ have both a prior ESS equal to 16. In contrast, their ECSSs are equal to 30 and 43 at $k = 50$. This suggests that in expectation 30 samples are added to, and 43 samples are subtracted from, the likelihood by the nonconflicting and conflicting prior, respectively, when comparing the MSEs of their posterior means to that of a baseline prior model with 50 observations.

In a real-data application, and particularly for small sample sizes, replacing $\theta_0$ by its (consistent) estimator introduces some bias and sample variability into the ECSS. Such error is derived and graphically explored for normal i.i.d. observations in Supporting Information Appendix A. A certain amount of error is common to all data-dependent approaches, and does not arise if the ECSS is used prospectively to quantify the impact of a prior for a range of hypothetical true values (see also Section 3).

The equivalence between the ECSS and prior ESS when no prior-data conflict is present depends on the chosen measure and on how conflict is defined. If the curvature as measure of information and the $\epsilon$-information prior are chosen, and $\theta_\pi = \theta_0$, the prior ESS of Morita et al. (2008) is retrieved. It can also be retrieved if, as for the normal distribution, the parameter controlling the curvature is independent from the parameter controlling the location of the posterior. If the MSE is chosen as a measure of informativeness, the normal-normal case provides a relevant example of the ECSS behavior (holding approximately for all other cases in which historical and current sample sizes are large). Under a dispersed baseline prior, the nominal prior ESS $\sigma_\pi^2 / \sigma_{\pi_0}^2$ is retrieved when $|\theta_\pi - \theta_0| / \sigma_\pi = 1$ (Neuenschwander et al., 2010). For values further away from the truth, the ECSS would be lower than the nominal prior ESS, while for values closer to the truth it would be larger. The latter is termed a “super-information” phenomenon by Reimherr et al. (2014), the approach of which leads to the same relationship as Neuenschwander et al. (2010) (although the targeted measure is partly different). For further details, we refer to Reimherr et al. (2014) and Neuenschwander et al. (2010).

### 3 | EFFECTIVE SAMPLE SIZES FOR ROBUST PRIORS

#### 3.1 | Robust mixture priors

The difference between prior ESS and ECSS becomes more subtle when considering priors designed to be robust with respect to prior-data conflict. A robust mixture prior $\pi(\theta) = (1 - \rho)\pi_0 \theta + \rho \pi_\theta(\theta)$, for $\rho \in [0, 1]$ representing the degree of skepticism, is a mixture of an informative and a baseline prior distribution resulting in a marginally heavy tailed distribution. This leads the prior information to be discarded in case of severe prior-data conflict (see, eg, Schmidli et al., 2014 and references therein).

As anticipated in Section 2.1, computing the prior ESS of a mixture prior is a controversial issue. The algorithm of Morita et al. (2008) can result in an unrealistically high

---

**FIGURE 2** MSE of posterior given normal data under nonconflicting prior (dashed), conflicting prior (dotted), and baseline prior (solid). ECSS at $k = 50$ is obtained by matching MSE of prior of interest and baseline prior. ECSS, effective current sample size; MSE, mean square error.
prior ESS as it only depends on the curvature at the prior mean, which would generally not be significantly influenced by the baseline component. Moreover, if the components’ modes have different locations, resulting in multiple peaks or in a highly skewed distribution, even more unrealistic results are obtained. A first possibility is to evaluate the curvature at the prior main mode rather than at the mean, as in Schmidli et al. (2014). A second approach is to approximate the mixture by a one component distribution (eg, normal) such that closed-form results are available. Of course, if this approximation does not fit well, the result may not be informative. A third approach is to compute the prior ESS for each mixture component and weight them according to the mixture weights, an approach also pursued by Gravestock and Held (2019). This is consistent with the concept that the prior arises from a historical sample—in this case from a heterogeneous population described by the mixture components—and it is thus the most coherent approach from our point of view. This approach does not pose any technical difficulty as the prior ESS of each component is well defined and available in closed form for the most common distributional choices. Note that the three approaches may yield different results. Further, they do not describe how much information the prior introduces for given data, which would only be captured by a data-dependent measure of the effective sample size. Therefore, it has been suggested to compute an effective sample size of the posterior and subtract the data sample size. Hobbs et al. (2013) (see also Neuenschwander et al., 2010) approximates the effective sample size of the posterior by a linear function of the ratio of posterior precisions under the prior of interest and under a reference prior, effectively assuming approximate normality. They call the resulting measure “effective historical sample size (EHSS)” and we adopt this term in the following. The approach of Schmidli et al. (2014) obtains the effective sample size of the posterior by applying to it the approach of Morita et al. (2008) for the prior ESS. We denote this approach “EHSS (MTM).” However, this solution may again be unsatisfactory when the posterior exhibits strong multimodality, as exemplified in Supporting Information Appendix B. Again, a weighted average of the prior ESSs of each posterior component helps overcoming this problem; we denote the latter approach “EHSS (mix).”

We compare the latter two ESS measures with the proposed data-dependent ECSS for varying degrees of prior-data conflict. Figure 3 shows the result for a normal-normal and a beta-binomial model. Such plot can be used to prospectively investigate the prior impact before seeing data. Specifically, for the normal-normal model we assume i.i.d. observations $y_i \sim N(\mu_0, 1)$, $i = 1, ..., n, n = 200$, for varying $\mu_0$, and assign to $\mu_0$ a baseline prior $N(0, 10^2)$, an informative prior $N(0, 1/50)$, or mixture prior $0.5N(0, 1/50) + 0.5N(0, 10^2)$. For the beta-binomial model we adopt the same prior and data specifications as Schmidli et al. (2014), that is, we assume data $y \sim Bin(n, p_0)$, $n = 40$, for varying $p_0$, and assign to $p_0$ a uniform baseline prior, an informative prior $Beta(4, 16)$, or mixture prior $0.5Beta(4, 16) + 0.5Beta(1, 1)$. No sampling variability is introduced in the computed ESS measures in this illustration, that is, the truth is assumed to be known exactly. The impact of sampling will be illustrated in Section 4. By construction, the data-independent prior ESS is constant across all data realizations (not shown): in the normal-normal model, the weighted average of the mixture component prior ESSs gives a value of 25, while the algorithm of Morita et al. (2008) provides an effective sample size of 49; in the beta-binomial model, coincidentally, both algorithms result in a prior ESS of 11. All data-dependent measures shown in Figure 3 approach the prior ESS of the baseline prior when conflict is severe. The ESS measures are also contrasted with the MSE induced by the mixture and the baseline prior, depicted below in dashed black and solid gray, respectively. The ECSS exactly captures the trend of the MSE: the MSE is decreased for no or very low conflict, as compared to the baseline prior model, that is, incorporating prior information is beneficial and the ECSS is positive. For moderate conflict, however, the MSE is increased and the ECSS gets negative. Here, the mixture prior does not fully discard the prior despite the conflict, and both mixture components obtain a relevant weight. In contrast to the ECSS, the “EHSS (mix)” is always positive while the “EHSS (MTM)” becomes negative but not consistently with the MSE behavior.

### 3.2 Empirical Bayes power and commensurate priors

Power and commensurate priors are additional robust prior specifications which allow discounting part of the available historical information. The power prior approach achieves this by viewing the available prior as a posterior from a historical trial, and thus proportional to the product of the historical data likelihood and a baseline prior. The likelihood of the historical data is then raised to a power $\delta \in [0, 1]$, which represents the proportion of the $n_h$ historical samples accounted for in the EHSS (eg, De Santis, 2007), that is, the EHSS is equal to $\delta n_h$. The choice or estimation of the discounting factor is commonly based on sensitivity analyses (eg, De Santis, 2007), full Bayes (eg, Neuenschwander et al., 2009), and empirical Bayes (Gravestock and Held, 2017). Here, we focus on the empirical Bayes power prior of Gravestock and Held (2017) (EB power prior). In the latter, $\delta$ is estimated by
taking into account the current observations through maximization of the marginal likelihood, thus giving rise to a data-adaptive approach. Note that in case of full Bayes estimation, it has been proposed to calculate the EHSS via approximation of the posterior by a mixture distribution or by using the marginal posterior mean of $\delta$ (Gravestock and Held, 2017; 2019) which may yield different results, an ambiguity which would not be present in the ECSS.

The commensurate prior approach (Hobbs et al., 2012) allows a distinction between the historical and current data parameter. The prior is defined as the product of the historical data likelihood $f_{\theta_0}(y_{1:n_0}|\theta_0)$, the historical parameter (vague) prior $\pi_{\theta_0}(\theta_0)$, and the commensurate prior $\pi(\theta|\theta_0, \eta)$, where $\eta$ is a commensurability parameter which can be fixed a priori, or estimated via full or empirical Bayes, as for the power prior approach. As for the power prior, we focus on the EB estimation of $\eta$, but the methodology is also applicable to the full Bayes counterparts. Note that for the normal-normal case, the EB power, and commensurate prior coincide (see Supporting Information Appendix C.1). For the binomial case, numerical procedures are applied to estimate $\eta$ via maximization of the marginal likelihood, and the EHSS is obtained by approximating the marginal posterior for $\theta$ by a Beta distribution, summing its parameters, and subtracting the current sample size.

Figure 3 includes a comparison of the different ESS measures for the EB power prior specification, and varying degrees of prior-data conflict. Left column, normal-normal model; right column, beta-binomial model. The true data-generating parameter values are assumed to be known. In the normal-normal model, curves are symmetric about the prior mean at $\mu_0 = 0$ and EB power prior results are equivalent to EB commensurate results. In the beta-binomial model, the ECSS is truncated at $n = 40$, that is, the total final sample size. Note that the use of a uniform baseline prior in case of a binary outcome leads to an ECSS converging to zero for extreme conflict while the EHSS converges to an EHSS of two (ie, the prior ESS of the uniform). Values of “informative (ECSS)” are truncated from below. ECSS, effective current sample size; ESS, effective sample size; MSE, mean square error [This figure appears in color in the electronic version of this article, and any mention of color refers to that version]
results of Figure 3 directly apply. Figure S7 contrasts the two EB priors for the beta-binomial model showing that the EB commensurate EHSS and ECSS tend to be right-shifted, and the ECSS is more unstable close to the boundaries, as compared to the EB power prior. The ECSS captures the fact that the use of the EB power (and commensurate) priors may still result in a minor conflict in terms of MSE. Moreover, the ECSS can be used as an intuitive tool to compare mixture, power and commensurate priors suggesting that, in the assumed setup, the mixture prior would allow for a larger prior-data conflict before discarding historical information. A different choice of the mixture weight $\rho$ and of the variance of the baseline mixture component could have, however, led to a different conclusion.

4 | ADJUSTING THE CONTROL SAMPLE SIZE IN AN ADAPTIVE TRIAL DESIGN

We illustrate the utility of the ECSS by applying it to a two-arm Bayesian clinical trial design where it is assumed that prior information on the control arm exists. In this situation, it is desirable to make use of this prior information to reduce the final sample size in the control group, while keeping operating characteristics close to an analysis with full sample size and without prior information. At the same time, robustness with respect to deviations from prior beliefs is aimed for. With this in mind, the following adaptive randomization scheme has been proposed for designs which aim at adapting the final sample size in the control arm according to the information contained in the prior distribution, as measured at an interim analysis. Variants of this design have been studied by several authors (eg, Hobbs et al., 2013; Schmidli et al., 2014; Bennett, 2018). All of them consider some version of the data-dependent EHSS as a measure of prior informativeness with priors which adapt to prior-data conflict. In this context, we argue that the ECSS is intuitively more appropriate than the EHSS as it answers the question of how many control samples are offset by the prior after stage two, that is, how many control patients are added or subtracted from the analysis by inclusion of prior information. In contrast, the prior ESS is independent of the currently observed control samples, while the data-dependent EHSS depends on the currently observed control samples according to a measure which may be problematic in situations of moderate conflict and is not among the trial’s targets, as outlined in Section 3.1.

To see this, we simulate a two-arm trial assuming normal outcomes $y_{\text{control}} \sim N(\mu_0, 1)$ and $y_{\text{treat}} \sim N(\mu_0 + \tau, 1)$ for varying control means $\mu_0$ and effect sizes $\tau \in [0, 0.28]$ performing $10^4$ Monte Carlo replications. Priors are defined as in Section 3 and variances are assumed to be known. Again, note that results for the EB commensurate prior are equivalent to those of the EB power prior and we only refer to the latter in the following. We conduct an interim analysis after 100 samples in each arm have been collected and compute the ESS measures in the control arm according to the different approaches. Then, additional 100 patients are recruited in the treatment arm while in the control arm we subtract from the target stage-two sample size of 100 a number of patients equal to the ESS measures. We contrast the ECSS with the prior ESS and EHSS for priors which do not and do adapt to prior-data conflict, respectively. In case of mixture priors, we consider the “EHSS (MTM)” as implemented by Schmidli et al. (2014), as well as the “EHSS (mix).” Note that the ECSS and the “EHSS (MTM)” can become negative in case of prior-data conflict, which would require an increase in the number of recruited patients above that required by a baseline prior model, which is not desirable in this context. Therefore, we replace the informative prior by a baseline prior (which is the limiting prior in case of extreme prior-data conflict for data-adaptive priors) if $ESS < 0$, effectively adding a second layer of robustness. Also note that we deviate from Schmidli et al. (2014) in that we project the observed data to the target sample size of $n = 200$ to estimate the EHSS and ECSS.

At the end of the trial, we evaluate the MSE of $\tau$ and frequentist type I error and power for hypotheses $H_0$: $\tau \leq 0$ vs $H_1$: $\tau > 0$, as based on the posterior probability with decision threshold $c$, $P(\tau > 0|y_{\text{treat}}, y_{\text{control}}) > c$. We consider a common fixed decision threshold $c = 0.975$. Thus, using the baseline prior, frequentist type I error is controlled at 2.5% and power for $\tau = 0.28$ is 80%.

Figure 4 shows the average ESS measures and the expected final sample size for the above design (see Figures S8 and S9 for their distributions). It demonstrates that the reduction in patient sample size is stronger using the ECSS as compared to the EHSS for control means close to the prior mean while more patients are needed for control means moderately far from the prior mean (moderate prior-data conflict). For large deviations from the prior mean, differences diminish. Figure 5 shows that all frequentist operating characteristics using the estimated ECSS are closer to those obtained under the baseline prior model, as compared to the ones obtained using any version of the estimated EHSS. We can see that the use of the ECSS avoids excessive inflation of type I error when the control mean is positive (prior shifts posterior of control downwards), and excessive power.
loss when the control mean is negative, as compared to the other approaches. The general trend in the behavior of the operating characteristics is in line with Viele et al. (2014), that is, improvements are observed when the prior information is consistent with the true parameter value generating the data. The remaining differences in operating characteristics under the baseline prior model when using the ECSS are explained by the fact that $\mu_0$ is replaced by an estimate, which yields a biased estimate of the ECSS and consequently MSEs differing from the targeted MSE under the baseline prior model, as outlined in Section 2.2. It has to be noted, however, that a certain amount of error has to be expected for any data-dependent approach, as illustrated in Supporting Information Appendix A. Figure S10 shows that equal MSEs as compared to the baseline prior model would indeed be obtained for all prior models under reduced sample sizes if we calculate the ECSS under a hypothetically known $\mu_0$ (shown in Figure 3), while the use of the EHSSs lead to different results. Thus, compared to the data-dependent EHSS approaches, the ECSS provides the advantage of being explicit in the targeted measure and, as shown in Figure S10, to get closer to the optimal discounting mechanism which we would adopt if interim data would provide perfect information. Note, however, that if strict control of type I error is required, no power gains could be achieved as compared to the design which adopts the baseline prior and the full sample size (Kopp-Schneider et al., 2019). In this sense, we stress that the ECSS should be interpreted as a tool to detect potential conflict and to robustify a design, but does not replace a careful elicitation of the prior distribution, in which a certain degree of trust has to be placed.

We can observe that the results obtained using the specified mixture prior and the informative prior, combined with prior discarding for negative ECSS, are quite close. The similarity depends on the specific choice of the mixture weights and variance of the baseline component. However, the results suggest that the ECSS may be used to obtain a robust design without the need to perform such a choice. Replacing the baseline component in the mixture prior by a unit information prior (Kass and Wasserman, 1996) would have yielded a similar result to the power/commensurate prior in the given setting (Figures S11 and S12).

We have additionally investigated different variations of the proposed design. Figures S11 and S12 shows the operating characteristics obtained if instead of discarding the prior for negative ECSS, the corresponding number of samples is added to the final sample size. The results show that for low degrees of prior-data conflict operating characteristics are slightly improved, while for larger degrees of prior-data conflict they are slightly worse. Further, we have investigated the effects of replacing the assumed control and treatment data variance by their empirical standard deviations in each Monte Carlo sample, observing that average results are almost...
identical, while as one would expect variability of results increases slightly for all priors and ESS approaches. Finally, we have replicated the design of Schmidli et al. (2014) with a binomial outcome. Similar operating characteristics are obtained for the ECSS and the “EHSS (MTM)” used by Schmidli et al. (2014). This is consistent with the similar ESS estimates provided by the two measures in this scenario, as shown in Figure 3.

5 | DISCUSSION

In this work, we propose to distinguish two frameworks of effective sample sizes of priors, prior ESS and ECSS, quantifying prior information in terms of historical (or external) and current (newly collected) samples, respectively. The former is suitable when a data-independent measure is desirable while the latter may be of interest when a data-dependent measure is desired. While the prior ESS quantifies the amount of information contained in a prior in terms of historical samples, the ECSS intends to additionally quantify its impact. For example, in a pediatric trial where the prior distribution is elicited from a preceding adult trial, we may both be interested in how many (hypothetical) patients with adult characteristics (as measured by prior ESS) or with child characteristics (as measured by ECSS) are added to the data-set of children through incorporation of the prior information. Both measures are valuable tools to provide an intuitive and easy to communicate quantification of the impact of a prior on posterior inference. It seems to us that there is a need to be specific about which ESS
perspective one takes and that, depending on the context, each of the measures prior ESS and ECSS have its place and can provide complementary information. The prior ESS can aid prior elicitation when no information about the future trial is yet available. In contrast, the ECSS could be useful to assess and communicate the impact of the prior on posterior inference and to detect and quantify prior-data conflict given simulated or real data. The adaptive design in Section 4 illustrates that there are situations in which the interest is explicitly on a data-dependent ECSS measure.

It has been noticed that no entirely convincing approach to ESS for the increasingly popular robust mixture priors exists to date (Neuenschwander and Schmidli, 2019). In this context, the ECSS provides a useful alternative which is able to show how the mixture prior adapts to the data, and to detect where possible residual prior-data conflict remains. As the likelihood for the current data is assumed to arise from a homogeneous population, it avoids the difficulty of quantifying the effective sample size of a mixture distribution. Such difficulty is not encountered for empirical Bayes power priors, which are typically assumed to be nonmixture posteriors arising from a given historical sample, directly discounted according to conflict. However, the EB power prior approach requires maximization of the marginal likelihood, which is easily available only for a restricted class of parametric models.

The ECSS is neither restricted to conjugate priors nor to the target MSE measure, although the latter is certainly an important target of interest in the clinical trial design context considered, and, more broadly, in an estimation framework. The flexibility in the target measure can be considered as an advantage as it allows to be explicit about the analysis objectives, and alternative measures such as type I/II error can be considered. We note, however, that the latter pose additional difficulties due to the unsmoothness in their decrease for increasing sample size, and further research may be needed.

Subsequent research will also include the quantification of the effective sample size in terms of current samples of hierarchical models (see Neuenschwander et al., 2010; Morita et al., 2012, for approaches to ESS quantification in this context), where the impact of prior distributions is particularly challenging to evaluate and a data-dependent measure may provide valuable insights into the intrinsic bias-variance trade-off. Note that the ECSS of meta-analytic-predictive priors can readily be computed by first approximating them by a mixture prior (Schmidli et al., 2014).

Clinical trials with adaptive adjustment of the control arm sample size represent a relevant applied context in which the ECSS measure can be of interest, as illustrated in Section 4. The ECSS allows to compute the number of control patients accounted for by the prior in an adaptive way, gradually decreasing as evidence of possible conflict is provided by the observed data. Being data-dependent, it is subjected to error. However, we show that it leads to a final sample size reduction which is closer than other ESS approaches to the sample size reduction we would apply if the truth was known (and if MSE represents the targeted measure). In this sense, it can be used as a stand-alone method to robustify a design which adopts an informative prior, or as an additional layer of robustness if, for example, mixture or EB power/commensurate priors are used.

ACKNOWLEDGEMENT

We thank Isaac Gravestock for a discussion on the EHSS for mixture priors as well as the AE and an anonymous referee for their insightful comments.

ORCID

Manuel Wiesenfarth ORCID: https://doi.org/10.17863/CAM.18107. (Accessed September 5, 2019).

REFERENCES

Bennett, M.S. (2018) Improving the efficiency of clinical trial designs by using historical control data or adding a treatment arm to an ongoing trial. PhD thesis, University of Cambridge. https://doi.org/10.17863/CAM.18107. (Accessed September 5, 2019).

Berger, J.O. and Berliner, L.M. (1986) Robust bayes and empirical bayes analysis with ϵ-contaminated priors. The Annals of Statistics, 14, 461–486.

Berger, J.O., Bernardo, J.M. and Sun, D. (2009) The formal definition of reference priors. The Annals of Statistics, 37, 905–938.

Bousquet, N. (2008) Diagnostics of prior-data conflict. The Annals of Statistics, 35, 1011–1029.

Box, G.E.P. (1980) Sampling and bayes’ inference in scientific modelling and robustness. Journal of the Royal Statistical Society, Series A, 143, 383–430.

Brueckner, M., Titman, A., Jaki, T., Rojek, A. and Horby, P. (2018) Performance of different clinical trial designs to evaluate treatments during an epidemic. PLOS One, 13, 1–19.

Chen, N., Carlin, B.P. and Hobbs, B.P. (2018) Web-based statistical tools for the analysis and design of clinical trials that incorporate historical controls. Computational Statistics and Data Analysis, 127, 50–68.

De Santis, F. (2007) Using historical data for bayesian sample size determination. Journal of the Royal Statistical Society, Series A, 170, 95–113.

Evans, M. and Moshonov, H. (2006) Checking for prior-data conflict. Bayesian Analysis, 1, 893–914.

Gamalo-Siebers, M., Savic, J., Basu, C., Zhao, X., Gopalakrishnan, M., Gao, A. et al. (2017) Statistical modeling for bayesian
extrapolation of adult clinical trial information in pediatric drug evaluation. *Pharmaceutical Statistics*, 16, 232–249.

Gelman, A., Meng, X.-L. and Stern, H. (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6, 733–760.

Gelman, A., Simpson, D. and Betancourt, M. (2017) The prior can often only be understood in the context of the likelihood. *Entropy*, 19, 555.

Gravestock, I. and Held, L. (2017) Adaptive power priors with empirical bayes for clinical trials. *Pharmaceutical Statistics*, 16, 349–360.

Gravestock, I. and Held, L. (2019). Power priors based on multiple historical studies for binary outcomes. *Biometrical Journal*, 61, 1201–1218.

Hobbs, B.P., Carlin, B.P. and Sargent, D.J. (2013) Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials*, 10, 430–440.

Hobbs, B.P., Sargent, D.J. and Carlin, B.P. (2012) Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis*, 7, 639.

Ibrahim, J.G., Chen, M.-H., Gwon, Y. and Chen, F. (2015) The power prior: theory and applications. *Statistics in Medicine*, 34, 3724–3749.

Kass, R.E. and Wasserman, L. (1996) The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91, 1343–1370.

Kopp-Schneider, A., Calderazzo, S. and Wiesenfarth, M. (2019) Power gains by using external information in clinical trials are typically not possible when requiring strict type I error control. *Biometrical Journal*. https://doi.org/10.1002/bimj.201800395

Morita, S., Thall, P.F. and Müller, P. (2008) Determining the effective sample size of a parametric prior. *Biometrics*, 64, 595–602.

Morita, S., Thall, P.F. and Müller, P. (2012) Prior effective sample size in conditionally independent hierarchical models. *Bayesian Analysis*, 7, 591–614.

Neuenschwander, B., Branson, M. and Spiegelhalter, D.J. (2009) A note on the power prior. *Statistics in Medicine*, 28, 3562–3566.

Neuenschwander, B., Capkun-Niggli, G., Branson, M. and Spiegelhalter, D.J. (2010) Summarizing historical information on controls in clinical trials. *Clinical Trials*, 7, 5–18.

Neuenschwander, B. and Schmidli, H. (2019) Use of historical data. In: Lesaffre, E., Baio, G. and Boulanger, B. (Eds.) *Bayesian Methods in Pharmaceutical Research*. Boca Raton: CRC Press.

Pocock, S.J. (1976) The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*, 29, 175–188.

Reimherr, M., Meng, X.-L. and Nicolae, D.L. (2014) Being an informed bayesian: Assessing prior informativeness and prior likelihood conflict. arXiv e-prints, arXiv:1406.5958v1. (Accessed September 5, 2019).

Schmidli, H., Gsteiger, S., Roychoudhury, S., O’Hagan, A., Spiegelhalter, D. and Neuenschwander, B. (2014) Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, 70, 1023–1032.

Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N. et al. (2014) Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics*, 13, 41–54.

Weber, S. (2018) *RBesT: R Bayesian evidence synthesis tools*. Available at: https://CRAN.R-project.org/package=RBesT. (Accessed 30 November, 2018).

**SUPPORTING INFORMATION**

Web Appendices and Figures referenced in Sections 3.2 and 4 are available with this paper at the Biometrics website on Wiley Online Library. An efficient implementation is provided in the R package “ESS” available on https://github.com/wiesenfa/ESS which extends package RBesT (Weber, 2018). It allows the computation of the prior ESS, EHSS and ECSS measures for binomial and normal outcomes. It contains a graphical tool to demonstrate the prior impact for a range of hypothetical true values and includes functions to reproduce the design in Section 4. Examples using it and description on how to reproduce Figures in Sections 3 and 4 are provided in Web Appendix E.

_How to cite this article:_ Wiesenfarth M, Calderazzo S. Quantification of prior impact in terms of effective current sample size. *Biometrics*. 2020;76:326–336. [https://doi.org/10.1111/biom.13124](https://doi.org/10.1111/biom.13124)