

Monitoring of sequential trials using a robust Bayesian stopping rule

Stefania Gubbiotti and Fulvio De Santis

Abstract In this paper we consider a method for monitoring a clinical trial whose patients are sequentially evaluated for response. We focus on a parameter representing treatment effect. Adopting a Bayesian approach we suggest to update progressively prior information on this unknown quantity: in particular, we monitor the trend of the posterior probability that the parameter is larger than a minimally clinical relevant value, as responses are collected. The trial is terminated with success as soon as this probability exceeds a prefixed threshold; if this does not happen before a preplanned maximum sample size is reached, the treatment is declared ineffective. Hence, the sample size is a random variable associated to the chosen stopping rule. With a simulation study we show that the expected sample size is always smaller than the preplanned optimal sample size and we illustrate an application to compare the sequential and the non sequential procedure. Finally, a robust version of the sequential criterion is proposed in which a single prior distribution is replaced by a suitable class of prior distributions.

Key words: Bayesian statistics, Design of experiments, Robustness, Sample size, Sequential analysis.

1 Introduction

In a sequential study the sample size is not planned in advance and a stopping rule is defined in order to terminate the trial as soon as significant results are observed. The advantage of sequential procedures is that they require on average a smaller number

Stefania Gubbiotti, e-mail: stefania.gubbiotti@uniroma1.it
Fulvio De Santis, e-mail: fulvio.desantis@uniroma1.it
Dipartimento di Statistica, Probabilità e Statistiche Applicate, Sapienza Università di Roma,
Piazzale Aldo Moro n. 5, 00185 Roma

of patients with respect to non sequential criteria. For references on this topic see among others [3, 5, 6, 7, 8, 9, 10].

In this work we consider a Bayesian sequential method for monitoring a trial and we compare the sequential expected sample sizes with the non sequential ones in a simulation study. Furthermore, we extend to the specific setting of sequential trials, the idea of allowing for additional uncertainty in the prior specification, introduced by [4] in the case of non sequential trials, and further developed in [1] and [2], where classes of ε -contamination priors and mixtures of prior distributions are respectively considered. Hence, we also introduce sequential robust criteria for the choice of the sample size and we compare them to the non robust ones.

The outline of this paper is as follows. In Section 2 we describe the general set up and we introduce some notation. In Section 3 we provide details of the sequential method for sample size selection and we derive a robust version of the criterion in Section 3.1. Section 4 is dedicated to an application. Finally Section 5 contains some concluding remarks.

2 Problem settings

Let us consider a phase II trial with the objective of estimating the effect of a new experimental treatment. The parameter of interest θ represents a continuous measure of treatment efficacy; for instance, in a superiority trial, the larger θ the more effective the treatment. We assume that groups of patients are sequentially evaluated for response and that the trial is terminated according to a given stopping rule (see Section 3). Hence, instead of prefixing a generic sample size we have to consider an increasing number of individuals $n_{[j]}$ where $j = 1, \dots, J$ is the group index and we set $n_{[j]}$ equal to a maximum total sample size N_{max} that is preplanned at the beginning of the trial. Notice that $n_{[j+1]} = n_{[j]} + k_j$, where $k_j \geq 1$ represents the size of the j -th group. Without loss of generality in the following we assume $k_j = 1$ for all $j = 1, \dots, J$, which means we consider each single patient sequentially.

For the generic j -th group of patients let us denote by $Y_{n_{[j]}}$ a measure of treatment response that is supposed to be normally distributed with mean θ and variance $\sigma^2/n_{[j]}$, with a prefixed value for σ^2 . Then, let $y_{n_{[j]}}$ and $f(y_{n_{[j]}}; \theta)$ denote the observed data and the corresponding likelihood respectively, for $j = 1, \dots, J$. Although in this case we are assuming normality, it is worth to remark that the same basic model provides an approximation that can be used, for instance, for binary data (with θ denoting a log-odds function) and for survival data (with θ denoting a log-hazard function), as discussed in [8].

Furthermore, in order to formalize pre-experimental knowledge on the phenomenon of interest consider a conjugate prior distribution on θ , that is a normal density of mean θ_0 and variance σ^2/n_0 , with known σ^2 and prior sample size n_0 , according to the notation of [8]. From Bayes theorem we know that the posterior distribution of θ given $y_{n_{[j]}}$ is $\pi_0(\theta|y_{n_{[j]}}) \propto f(y_{n_{[j]}}; \theta)\pi_0(\theta)$ and for the normal model with conjugate priors we have

$$\pi_0(\theta|y_{n[j]}) = N\left(\theta \middle| E_{n[j]}, V_{n[j]}\right), \quad (1)$$

where the posterior expectation and the posterior variance are respectively

$$E_{n[j]} = \frac{n_0\theta_0 + n[j]\bar{y}_{n[j]}}{n_0 + n[j]} \quad \text{and} \quad V_{n[j]} = \frac{\sigma^2}{n_0 + n[j]},$$

with $\bar{y}_{n[j]} = \frac{\sum_{i=1}^j y_{n[i]}}{n[j]}$ representing the sample average computed over the first $n[j]$ observed response values. Using (1) it is straightforward to update progressively information on θ as each value of the response $y_{n[j]}$ is observed, for $j = 1, \dots, J$, and to establish a stopping rule for the trial as proposed in the next section.

3 A sequential criterion for sample size selection

In this section we describe a sequential criterion that has already been introduced and discussed for instance in [8].

As posterior quantity of interest, let us consider the posterior probability that θ exceeds a minimally clinical relevant value δ :

$$P_{\pi_0, n[j]}(\theta > \delta|y_{n[j]}) = 1 - \Phi\left(\frac{\delta - E_{n[j]}}{\sqrt{V_{n[j]}}}\right). \quad (2)$$

The treatment is declared successful if experimental evidence that probability (2) is larger than a given threshold $\gamma \in (0, 1)$ is sufficiently strong. At the generic j -th iteration, we proceed as follows:

1. Enroll the j -th group of patients and observe $y_{n[j]}$
2. Update the posterior parameters $E_{n[j]}$ and $V_{n[j]}$
3. Compute the posterior probability (2)
4. *Stopping rule:* if

$$P_{\pi_0, n[j]}(\theta > \delta|y_{n[j]}) > \gamma \quad (3)$$

the trial *stops with success*, otherwise the procedure is repeated for the $(j+1)$ -th group of patients starting from step 1.

If condition (3) is never fulfilled and the maximum preplanned number of patients N_{max} is reached, the trial is terminated *without success*.

In this context the sample size is a random variable $\mathbb{N}$ whose distribution depends on the stopping rule given by (3). In Section 4 the distribution of $\mathbb{N}$ is simulated drawing the data sequentially from the sampling distribution $f(\cdot; \theta_D)$, where θ_D is a design target value for treatment effect (see Figure 2). As proposed in [4, 1, 2], it is possible to adopt a fully predictive approach by using, instead of the sampling

distribution, the marginal distribution of the data $m_D(\cdot) = N(\cdot | \theta_D, \sigma^2 (n^{-1} + n_D^{-1}))$ induced for instance by a normal design prior $\pi_D(\theta) = N(\theta | \theta_D, \sigma^2 / n_D)$. This would allow one to take into account the uncertainty involved in the choice of the design value θ_D . Here, we do not go into further details of this approach since it does not affect the general idea of sequential trials.

It is interesting to remark that the expected value of $\mathbb{N}$ conditional to θ_D , will be smaller than or equal to the optimal sample size obtained choosing the corresponding non-sequential criterion introduced in [4, 1], i.e. $\mathbb{E}(\mathbb{N}) \leq n^*$, where

$$n^* = \min\{n : \mathbb{E}[P_{\pi_0}(\theta > \delta | y_n)] > \gamma\}. \quad (4)$$

In this sense, adopting a sequential procedure allows one to save observations, as we illustrate in the application of Section 4.

3.1 Robust sequential criterion

As discussed for instance in [1], in order to take into account the uncertainty due to the prior specification, it is possible to consider a class of prior distributions Γ instead of a single prior π_0 . This allows us to derive a *robust* version of the sequential criterion of Section 3. In the previously described set up, since we reach success when the posterior probability (2) exceeds a threshold γ , we need to focus on the inferior bound of $P_{\pi, n_{[j]}}(\theta > \delta | y_{n_{[j]}})$ as the prior π varies into the class Γ . Thus, condition (3) becomes:

$$\inf_{\pi \in \Gamma} P_{\pi, n_{[j]}}(\theta > \delta | y_{n_{[j]}}) > \gamma, \quad (5)$$

for each $j = 1, \dots, J$. This robust sequential criterion yields sample sizes that are uniformly larger than those determined with the non robust sequential procedure. As noted in [1], the same relationship holds true for the non sequential criterion (4) with respect to its corresponding robust version (that is similarly obtained replacing the posterior probability in (4) by its inferior bound for $\pi \in \Gamma$), i.e. $n^* < n_F^*$, where n_F^* denotes the robust (non sequential) optimal sample size. Finally, if we define $\mathbb{N}_F$ the random variable associated to the stopping rule in (5), we have that $\mathbb{E}(\mathbb{N}_F) \leq n_F^*$.

4 Application

In this section, we compare the sequential criterion performance to the non-sequential one. For the robust version of the method, we consider the restricted conjugate class of prior distributions $\Gamma_{RC} = \{\pi : \pi(\theta) = N(\theta | \theta_0, \sigma^2 / \tilde{n}), \tilde{n} \in [n_L, n_U]\}$, although other options would be feasible (see for instance [1]).

As an example, let us consider a clinical trial aimed at assessing the efficacy of a drug against leukaemia. Let us suppose that treatment response consists in a complete molecular remission that is assessed in terms of the reduction of a specific transcript associated with the disease of concern: the larger the reduction the more effective the treatment. In this setting, θ is the logarithm of the reduction, which is reasonably assumed to be normally distributed. Furthermore, we set as a minimally clinical relevant reduction $\delta = 2$. In practice, in the following we consider a fictitious dataset of observed responses for 50 patients and we assume the data to be collected sequentially. Moreover, based on the results of previous studies, we elicit a normal prior distribution of parameters $\theta_0 = 1.5$, $\sigma^2 = 4$, $n_0 = 10$ and we set a threshold on the posterior probability scale equal to $\gamma = 0.9$.

Hence we can proceed as described in Section 3: the trial stops as soon as we have evidence that our objective is proved, otherwise we continue up to the maximum number of patients $N_{max} = 50$. The obtained results are represented in Figure 1: the posterior probability that $\theta > \delta$ (black circles) is sequentially updated until it exceeds the threshold γ , that is after the 23-th patient is examined. Since condition (3) is fulfilled the trial reaches success and is terminated. Adopting the robust version (5) of the sequential criterion, with $\Gamma = \Gamma_{RC}$ and $(n_L, n_U) = (3, 30)$, the required number of patients to satisfy the stopping rule increases to 47. Notice that in this case we are assuming to have an observed fixed dataset and we are mimicking the sequential accrual to perform the proposed procedure.

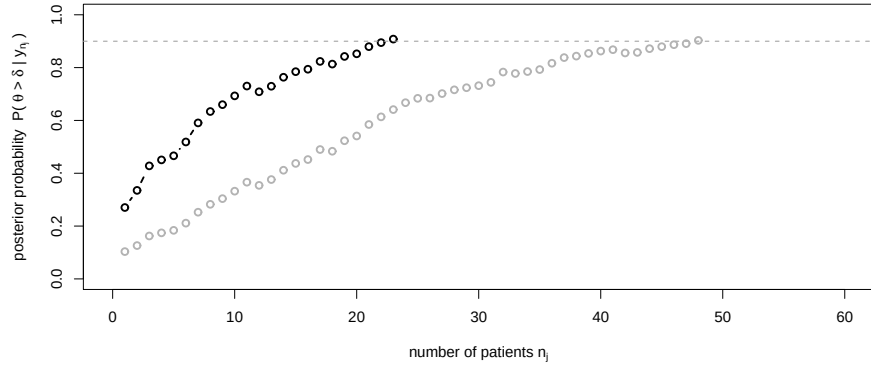


Fig. 1 Posterior probability $P_{\pi_0}(\theta > \delta | y_{n_{[j]}})$ (black circles) and inferior bound of the posterior probability $\inf_{\pi \in \Gamma} P_\pi(\theta > \delta | y_{n_{[j]}})$ (gray circles) for a sequentially increasing number of patients.

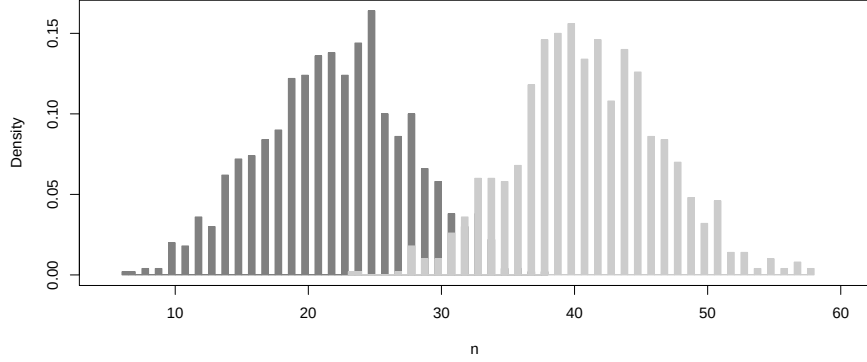


Fig. 2 Simulated distribution of $\mathbb{N}$ (dark grey) and $\mathbb{N}_\Gamma$ (light grey), choosing $\Gamma = \Gamma_{RC}$ with $(n_L, n_U) = (3, 30)$.

4.1 Simulation study

Conversely, if we simulate a large number M (for instance $M = 1000$) of datasets from the sampling distribution, conditional on a prefixed design value $\theta_D = 2.9$, we obtain M different “sequential trajectories” of the posterior probability: this yields M values for the sample sizes depending on the stopping rule (3) or its robust version (5). The simulated distributions of the random variables $\mathbb{N}$ (dark grey) and $\mathbb{N}_\Gamma$ (light grey) are represented in Figure 2. Their (approximated) expected values are $\mathbb{E}(\mathbb{N}) = 22$ and $\mathbb{E}(\mathbb{N}_\Gamma) = 41$, thus we have $\mathbb{E}(\mathbb{N}) < \mathbb{E}(\mathbb{N}_\Gamma)$ as argued in Section 3.1.

These values can be compared with the optimal sample sizes obtained adopting the non sequential (standard and robust respectively) criteria. The expected posterior probability $\mathbb{E}[P_{\pi_0}(\theta > \delta|y_n)]$ (black dots) involved in criterion (4) and $\mathbb{E}[\inf_{\pi \in \Gamma_{RC}} P_\pi(\theta > \delta|y_n)]$ (grey dots) are plotted as functions of n in Figure 3. Given the threshold $\gamma = 0.9$, as expected we have $n^* = 30 > \mathbb{E}(\mathbb{N}) = 22$ and $n_\Gamma^* = 50 > \mathbb{E}(\mathbb{N}_\Gamma) = 41$.

In Table 1 we compare sequential and non sequential sample sizes for several choices of the design value θ_D , using both standard and robust criteria. As the design value increases we uniformly get smaller sample sizes. Notice that, for instance, when $\theta_D = 2.5$ and Γ is moderately wide (for instance $n_U = 25$ and $n_U = 30$), the maximum number of patients is actually reached so that the sequential trial stops without success. Focusing on each pair of columns in the table we can confirm that the sequential procedure always allows one to save observations with respect to the non sequential sample size criterion. Moreover, when Γ is sufficiently small (for instance for $n_U = 15$) the robust sequential sample size $\mathbb{E}(\mathbb{N}_\Gamma)$ even entails an advantage in terms of observations saving with respect to the non robust and non sequential optimal sample size n^* , as shown comparing the third and the fourth

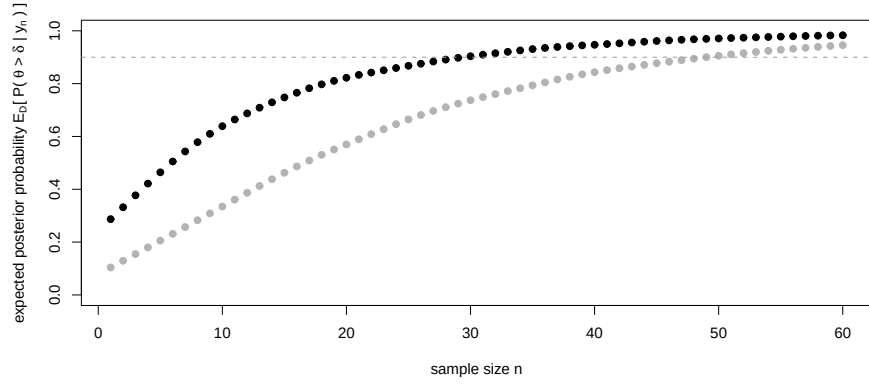


Fig. 3 Conditional expected posterior probability as a function of the sample size using both the non robust criterion (black dots) and the robust criterion (grey dots) respectively, with $\Gamma = \Gamma_{RC}$, $(n_L, n_U) = (3, 30)$.

Table 1 Sequential and non sequential sample sizes for several choices of the design value θ_D , using both the standard criteria ($n_0 = 10$) and the robust criteria with different classes of restricted conjugate prior distributions according to the choice of n_U .

θ_D	$n_0 = 10$		$n_U = 15$		$n_U = 20$		$n_U = 25$		$n_U = 30$	
	$\mathbb{E}(\mathbb{N})$	n^*	$\mathbb{E}(\mathbb{N}_\Gamma)$	n_Γ^*	$\mathbb{E}(\mathbb{N}_\Gamma)$	n_Γ^*	$\mathbb{E}(\mathbb{N}_\Gamma)$	n_Γ^*	$\mathbb{E}(\mathbb{N}_\Gamma)$	n_Γ^*
2.5	50	74	60	84	90	91	100	100	100	100
2.7	31	41	38	49	44	56	50	63	55	70
2.9	22	30	27	35	32	39	37	45	41	50
3.1	17	21	21	26	25	31	29	34	37	38
3.3	14	17	18	21	21	24	24	28	27	31
3.5	12	14	15	18	18	20	20	23	23	26

• $N_{max} = 100$

column in Table 1. Finally, looking at the robust criteria we notice that, as expected, wider classes of prior distributions (corresponding to increasing values of the upper bound for the prior sample size) yield larger values for the optimal sample size, regardless of the criterion being sequential or not.

5 Conclusions

In this paper we have shown how a sequential procedure allows early termination when there is evidence of treatment efficacy and enables the experimenter to reach a conclusion much earlier than in a typical study with fixed sample size. This is

very natural in a Bayesian context, since updating information on the parameter of interest as patients are enrolled, treated and evaluated for response, just translates in a sequential application of Bayes theorem and in a straightforward condition on a quantity of interest to be checked.

This work could be extended in several directions. First of all a predictive approach could be considered instead of a conditional one, as mentioned in Section 3, allowing one to take into account uncertainty involved in the choice of the design value (see [4, 1]). Furthermore, in this context one could also update the marginal distribution of the data in the light of the sequentially observed outcomes, as discussed in [2]. With respect to robustness, it would be possible to further investigate different alternatives for the classes of prior distributions, such as the ε -contamination classes adopted for instance in [1].

References

1. Brutti, P., De Santis, F., Gubbiotti, S.: Robust Bayesian sample size determination in clinical trials. *Stat. Med.* **27**, 2290-2306 (2008)
2. Brutti, P., De Santis, F., Gubbiotti, S.: Mixtures of prior distributions for predictive Bayesian sample size calculations in clinical trials. *Stat. Med.* **28**, 2185-2201 (2009)
3. Cui, L., Hung, H.M., Wang, S.J.: Modification of sample size in group sequential clinical trials. *Biometrics* **55**(3), 853-7 (1999)
4. De Santis, F.: Sample size determination for robust Bayesian analysis. *J. Am. Stat. Assoc.* **101**(473), 278-291 (2006)
5. Gao, P., Ware, J.H., Mehta, C.: Sample size re-estimation for adaptive sequential design in clinical trials. *J. Biopharm. Stat.* **18**(6), 1184-96 (2008)
6. Lehmacher, W., Wassmer, G.: Adaptive Sample Size Calculations in Group Sequential Trials. *Biometrics* **55**(4), 1286-1290 (1999)
7. Li, G., Shih, W.J., Xie, T., Lu, J.: A sample size adjustment procedure for clinical trials based on conditional power. *Biostatistics* **3**(2), 277-87 (2002)
8. Spiegelhalter, D. J., Abrams, K. R., Myles, J. P.: Bayesian approaches to clinical trials and health-care evaluation. Wiley (2004)
9. Shih, W.J.: Group sequential, sample size re-estimation and two-stage adaptive designs in clinical trials: A comparison. *Stat. Med.* **25**, 933-941 (2006)
10. Tang, L.L.: Sample size recalculation in sequential diagnostic trials. *Biostatistics* **11**(1), 151-163 (2010)