Contemporary Clinical Trials 29 (2008) 740–750

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Predictive checking for Bayesian interim analyses in clinical trials ☆

Takashi Daimon *

Department of Biostatistics, Hyogo College of Medicine, 1-1 Mukogawa-cho, Nishinomiya-city, Hyogo, Japan

ARTICLE INFO

Article history:
Received 31 January 2008
Accepted 13 May 2008

Keywords:
Prior predictive
Posterior predictive
Predictive probability
Model checking
Model criticism
Interim monitoring
Early stopping

ABSTRACT

Bayesian methodologies have been used for interim analyses of clinical trial data. In Bayesian interim analyses, decisions regarding the continuation of a trial are guided by a Bayesian model or indices, e.g., the predictive probability derived from it that specifies the conditions under which the clinical trial results might be judged sufficiently convincing to allow early stopping. Thus, its appropriateness for making such decisions depends on whether the model or the indices are reliable. In this paper we describe the use of both prior- and posterior- predictive checking approaches as a diagnostic tool for assessing the reliability of the model or indices on which the decision making is based. The proposed approach is illustrated with three examples, one of which is a simulation.

1. Introduction

For ethical, administrative and financial reasons, clinical trial data are often analyzed repeatedly at any time prior to the formal completion of a trial [1]. Such an analysis is called an interim analysis. A number of approaches have been developed for the interim analysis [2]. For example, the interim analysis is often implemented using the group sequential or stochastic curtailment approaches. The group sequential approach includes repeated significance tests [3] and a boundaries approach [4]. The concept of stochastic curtailment approach was proposed by Lan et al. [5] and Halperin et al. [6] and further developed by Betensky [7]. The idea here is to stop a trial as soon as an outcome of interest is determined with high probability. This approach makes use of the conditional (or interim) power function, with a study being more likely to be abandoned if its conditional power is poor. Applications of stochastic curtailment have been described by Anderson [8], Halperin et al. [9] and Hunsberger et al. [10] among others.

Increasingly, however, there has been much interest in the design and analysis of clinical trials and applications in health sciences under a Bayesian paradigm and many authors have recently discussed the role and implementation of Bayesian methods in the interim analysis of clinical trial data [3,11–15]. In particular, Spiegelhalter et al. [12] presented three categories of the Bayesian interim analysis, utilizing classical, hybrid and Bayesian predictions respectively. The interim analysis using classical prediction (corresponding to a likelihood approach) is a special case of the hybrid and Bayesian predictions below. This approach allows us to make predictions solely on the basis of the data obtained so far and provides one tool for those carrying out formal or informal interim analyses. However, in other words, we could say that the drawback of this approach is that it does not use the prior information. The interim analysis based on the hybrid prediction (corresponding to a mixed Bayesian-frequentist approach) could avoid this drawback. This approach uses a prior and data so far to predict a future frequentist analysis. Most Bayesian methodologies for the interim analysis have been based on the hybrid prediction (see e.g., [16–23]). However, Dmitrienko and Wang [24] emphasized "this approach has been criticized in the literature because it does not have a clear frequentist interpretation and, at the same time, is inconsistent with principles of Bayesian theory". In this study, we follow Dmitrienko and Wang's lead and focus on the Bayesian

☆ This research was supported partly by a grant from the Japanese Ministry of Education, Culture, Sports, Science and Technology (Grant number 18700272).
* Corresponding author. Tel./fax: +81 798 45 6169.
E-mail address: daimon@hyo-med.ac.jp.

1551-7144/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.cct.2008.05.005

prediction. The interim analysis using Bayesian prediction (corresponding to a fully Bayesian approach) includes a prior opinion for making predictions and in the analysis, and comprehends the classical predictions and the hybrid predictions (see e.g. [14] and Appendix A).

In Bayesian approaches one needs to construct the Bayesian model that provides the posterior information by combining the data with the prior. This holds true for Bayesian interim analyses. Naturally enough, decisions regarding the continuation of a trial are based on the model or indices derived from it, e.g., the predictive probability that specifies the conditions under which the clinical trial results might be judged sufficiently convincing to allow early stopping. The appropriateness of the judgment on early termination/continuation of a trial depends on whether the model or the indices are reliable or not.

Here we must acknowledge the insight of Box [25] that "all models are wrong, but some are useful". In other words, we may have to check critically to find out if the model can be considered to be generating the actual data. Then we can consider two useful approaches to such model checking or criticism: namely, the prior predictive checking approach [26] and the posterior predictive approach [27,28]. However it is noted that there is the difference in interpretation of the prediction between the two approaches (see Section 3). That is, the former provides checking models or indices by comparing data to the prior predictive distribution. This approach only contrasts information from the prior and data, and checks their compatibility. The latter does by comparing data to the posterior predictive distribution. This approach only contrasts information from the posterior and the data, and checks their compatibility. The prior predictive checking approach to clinical trials was illustrated by e.g., Spiegelhalter et al. [12,14], but no one have mentioned the application of the posterior predictive checking approach to interim analyses. Therefore, in order to make the full use of the ideas of both the prior- and posterior- predictive checking approaches and provide sufficient evidence on decisions regarding early termination/continuation of the trial, in this paper we propose one approach that uses both the prior predictive checking approach and the posterior predictive checking approach.

The predictive probability that enables us to conduct interim analyses using the Bayesian prediction is provided in Section 2. The prior- and posterior- predictive distributions are derived and the role and implementation of the predictive checking approach are introduced. Its performance is illustrated using three examples, one of which is a simulation, in Sections 4 and 5. Finally, Section 6 contains our concluding remarks.

2. Interim analyses using Bayesian prediction

To provide interim analyses using Bayesian prediction, consider a two-arm clinical trial comparing a test treatment to a standard treatment (either placebo or active control). The total projected sample size of the trial is n patients and the $k(=1,...,K)$th interim analysis is conducted after n_k patients have completed the trial (thus n_{K+1} $(=n)$ denotes the number of patients at the final analysis). To simplify presentation, we will assume throughout the paper that equal numbers of patients are enrolled in each treatment group. The Bayesian prediction approach discussed below is easily extended to the more general case of several treatment arms with unequal numbers of patients. Further, suppose that we are interested in predicting whether the future data will result in a posterior probability for the null hypothesis of no treatment effect H_0: $\theta<\delta$ against an alternative hypothesis H_A: $\theta\geq\delta$, being less than some small value ε. Here, θ denotes the parameter for a treatment difference of interest and δ represents a clinically significant improvement over the standard treatment. Throughout this paper $p(\cdot)$ and $\phi(\cdot)$ denote density functions.

We shall assume that our data after n_k observations can be summarized by a statistic y_{n_k}, whose distribution is

$$p(y_{n_k}|\theta) = \phi(y_{n_k}|\theta, \sigma^2/n_k), \tag{1}$$

where ϕ represents a normal distribution with mean θ and variance σ^2/n_k and σ^2 is assumed known: in two-arm comparative trials θ is the true treatment difference and y_{nk} is the parameter estimate. This assumption of a normal likelihood covers many situations [12]: if patients' responses are assumed normal with variance $\sigma^2/2$, θ is the true difference in mean response, and y_n is the difference in group sample means where n_k patients are allocated to each treatment; in survival analysis with proportional hazards, if there are $n_{T,k}$ events in the treatment group and $n_{C,k}$ events in the control group out of n_k total events in the kth interim analysis, then $y_{nk}=2(n_{T,k}-n_{C,k})/n_k$ has a distribution approximately given by Eq. (1) with mean θ and variance $4/n_k$, where θ is the log-hazard ratio [29]. For rare events, we have a similar approximation in which θ is the log-odds ratio, n_k is the number of events, y_{n_k} is the estimated log-odds ratio and $\sigma^2=4$. For binominal responses with higher event rates, y_{n_k} is the difference in sample response rates and, strictly speaking, σ^2 depends on the unknown response rates, but σ^2 may be used in Eq. (1) which for sufficiently large n_k will be adequate. Rate ratios can also be handled within this framework [30].

Let $p(\theta)$ denote the prior distribution. With a normal likelihood it is mathematically convenient, and often reasonably realistic, to make the assumption that $p(\theta)$ has the form

$$p(\theta) = \phi(\theta|\mu, \sigma^2/n_0), \tag{2}$$

where μ is the prior mean. This prior is equivalent to a normalized likelihood arising from a hypothetical trial of n_0 patients with an observed value μ of the treatment difference statistic (thus n_0 can be considered to be a special case of n_k, i.e., the number of patients in a hypothetical interim analysis before the trial starts). We shall make use of this normality assumption in the expressions shown below and in our examples.

Further, suppose that after we have observed n_k patients, we are interested in the possible consequences of continuing the trial for a further $n-n_k$ observations. In other words, suppose that we have observed a parameter estimate y_{n_k} based on sample size n_k at each interim analysis, and are considering a further $n-n_k$ observations which will yield a future parameter estimate $\tilde{y}_{n-n_k}$(it is noted that this estimate cannot be observed at each of the interim analyses and is assumed to have the distribution, $p(\tilde{y}_{n-n_k})=\phi(\tilde{y}_{n-n_k}|\theta, \sigma^2/(n-n_k))$). Then, for this assumed estimate $\tilde{y}_{n-n_k}$, we are interested in a 'significant' result S_ε^B which we have defined as