STATISTICS IN TRANSITION new series, December 2023
Vol. 24, No. 5, pp. 185–202, https://doi.org/10.59170/stattrans-2023-070
Received – 20.02.2023; accepted – 10.07.2023

Bayesian predictive probability design: theory and practical application in a prospective study

Adam Korczyński[1]

Abstract

In an experiment-based prospective study aiming to determine the efficiency of a treatment, the time by which it becomes clear whether a therapy is effective or not is critical. This applies specifically to clinical trials and refers to the same extent to both successful and futile therapies. This study seeks to answer the question when there is enough evidence allowing the trial to be finalised. The key is to find enough statistical signals, working on the smallest possible sample, to make a judgment whether to extend, continue or terminate the study. The Bayesian predictive design allows drawing conclusions about the prognosis of a study considering the actual results.

The article provides a theoretical background and presents a practical perspective, addressing the statistical properties and technical aspects of conducting a trial based on a predictive design. Additionally, the sensitivity of the design to the choice of prior distribution is discussed.

Key words: Prospective study analysis, adaptive design, predictive probability design, Bayesian statistics.

1. Bayesian adaptive design - overview

Bayesian predictive design[2] falls within the concept of adaptive design of a clinical trial allowing modifications of the trial conduct according to the intermediate results. One of the objectives is to end the trial as soon as the final outcome is apparent (George, Wang, & Pang, 2016, pp. 366–369). The predictive power within the clinical trial settings is described as "having a positive result from a trial based on the currently available data" (Heath, et al., 2020, p. 2). In fact, Bayesian design allows assessing positive or negative result and therefore the rules of stopping for either futility or efficacy.

[1] SGH Warsaw School of Economics, Poland. E-mail: akorczy@sgh.waw.pl. ORCID: https://orcid.org/0000-0002-1533-7197.

[2] An overview of the Bayesian approach including comparison with the frequentist principles can be found in (Lesaffre & Lawson, 2012).

Additionally, it gives the ability to react to evidence on superior treatment factor allowing assignment of new subjects to the best dosing scheme under study. For clinical trials various incentives, primarily ethical (Zhou, Liu, Kim, Herbst, & Lee, 2008, p. 2; Yin, Chen, & Lee, 2012, p. 220) but also those related to safety, cost and time effectiveness (Heath, et al., 2020; Chen, Ibrahim, Lam, Yu, & Zhang, 2011) bring the adaptive design approach into the scope of methods attractive for the industry and regulatory agencies. In some settings of oncology, the studies require large number of patients and years of studies in order to gain the approval (Barker, et al., 2009). Therefore, timing and decision making is essential to the drug development.

The Bayesian design finds its application in non-inferiority trials which are with the aim to show similar effectiveness of a novel therapy compared to a standard but with additional benefits to patients. If a therapy turns to be inferior then it would ideally be stopped early. Clinical practice shows that in oncology interim analysis was limited in recent years and based on a review published in 2012 (Heath, et al., 2020, p. 2), only 36% of 72 non-inferiority oncology trials utilized a formal analysis.

The Bayesian framework can also be applied to targeted therapy, giving the opportunity of accounting for the variation in patients characteristics (biomarker profile) when assigning the therapy. An application of targeted therapy in patients with advanced non-small cell lung cancer with disease control rate as the primary endpoint is outlined in (Zhou, Liu, Kim, Herbst, & Lee, 2008). The disease control rate was monitored within the treatment and marker subgroups with the Bayesian design deciding about the randomization of patients to treatment arms, according to the ongoing assessment of the response. A simulation study showed higher disease control rate in the randomized patients in the adaptive design as compared to fixed randomization equivalent (Zhou, Liu, Kim, Herbst, & Lee, 2008, p. 11; Yin, Chen, & Lee, 2012, p. 231). Similarly, application is to be found in (Barker, et al., 2009) with the description of a trial aiming at identifying the biomarker profiles able to predict the response for each study treatment.

Another application is within the basket design comparing the results of the treatment in patients with tumors of various types (Simon, Geyer, Subramanian, & Roychowdhury, 2016). The goal is to detect the strata in which drug activity suggests promising future results, and those which should not be continued as the evidence is the opposite. In randomized settings adaptive design allows assigning patients to more promising therapies, and such a schedule can already be applied to phase II trials (Yin, Chen, & Lee, 2012; Harrington & Parmigiani, 2016). The drawback is though that adaptive design in randomized setting would require larger sample due to unbalanced patient allocation (Yin, Chen, & Lee, 2012, p. 234).

With appropriate design of the trial the ability to determine the final outcome could be very high. An overview of the sensitivity of the Bayesian predictive probability design

for non-informative priors is provided in (Mitchell, 2018). The conclusion from the study was that in about 93% simulated trials, the interim decision was in agreement with the final outcome.

Bayesian design offers a straightforward setting for implementing prior knowledge into the estimation. The gain from using the prior knowledge is in the possibility to reduce the final sample size ensuring sufficient power to detect the effect (Chen, Ibrahim, Lam, Yu, & Zhang, 2011, p. 1167). In the referenced study, the reduction of the sample size between an informative and non-informative approach was from n=1480 to 1080 patients. This finding would not be however applicable to randomized trials with adaptive design, as then the distribution of patients over the treatment arm is affected by the outcome and this introduces the lack of balance between the arms.

The following sections describe and discuss the special case of a Bayesian design with predictive probabilities at the end of the trial using a binary endpoint. An example application is provided for a response rate endpoint.

2. Theoretical background of Bayesian predictive probability for binary outcome

The parameter of interest is the number of responses which can be translated into a response rate π. The objective is to predict the final response rate observed when the maximum sample size is reached, based on the initial assumption on the distribution of potential response rates and the actual outcome at a given time point. Within Bayesian settings it requires defining the prior distribution $f(\theta)$, which reflects the primary expectation on the response rate and the likelihood $L(y_1, y_2, \ldots, y_n| \theta)$, which adjusts the initial belief by the empirical evidence. The posterior distribution of the response rate is derived using the Bayesian rule (Lesaffre & Lawson, 2012, p. 23):

$$f(\theta \mid y_1, y_2, \ldots, y_n) = \frac{L(y_1, y_2, \ldots, y_n \mid \theta) f(\theta)}{\int L(y_1, y_2, \ldots, y_n \mid \theta) f(\theta) d\theta}. \tag{1}$$

The response rate can be seen as a probability of an individual success in a Bernoulli experiment $\pi=P(Y=1)$. The likelihood function for a sample of Bernoulli random variables $y_1, y_2, \ldots, y_n$ is given by the formula (Jóźwiak & Podgórski, 2009, pp. 193-4.):

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \mid \pi) &= \pi^{y_1}(1-\pi)^{1-y_1}\pi^{y_2}(1-\pi)^{1-y_2}\ldots\pi^{y_n}(1-\pi)^{1-y_n} \\ &= \pi^{\sum_i y_i}(1-\pi)^{n-\sum_i y_i}. \end{aligned} \tag{2}$$

It follows that the number of responses $R = \sum_{i=1}^{n} y_i$ is a random variable with a binomial distribution defined by the probability (Lesaffre & Lawson, 2012, p. 25):

$$P(R=r \mid \pi)=\binom{n}{r}\pi^{r}(1-\pi)^{n-r} \quad for\ r=1,2,\ldots,n. \tag{3}$$

The initial belief on the response rate can be expressed by the beta distribution, which has the following probability density function (Bolstad, 2007, p. 127):

$$f(\pi \mid a,b)=\begin{cases}\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1} & for\ \pi\in[0,1],\\ 0 & for\ \pi\notin[0,1]\end{cases} \tag{4}$$

where $\Gamma(s)=\int_0^\infty t^{s-1}e^{-t}dt$, s>0 is the Gamma function, and a>0 and b>0.

The beta distribution associates probability with any value between 0 and 1 which covers all possible response rates. If there is no strong evidence for any particular set of values one can use Beta prior with parameters a=0.5 and b=0.5, which gives similar probability for the values in the central region of π distribution and with somewhat more probable observation of extreme values (see Figure 1). Alternative formulation with a=1 and b=1 would result in uniform distribution assuming the same probability for all the possible values of the parameter of interest.

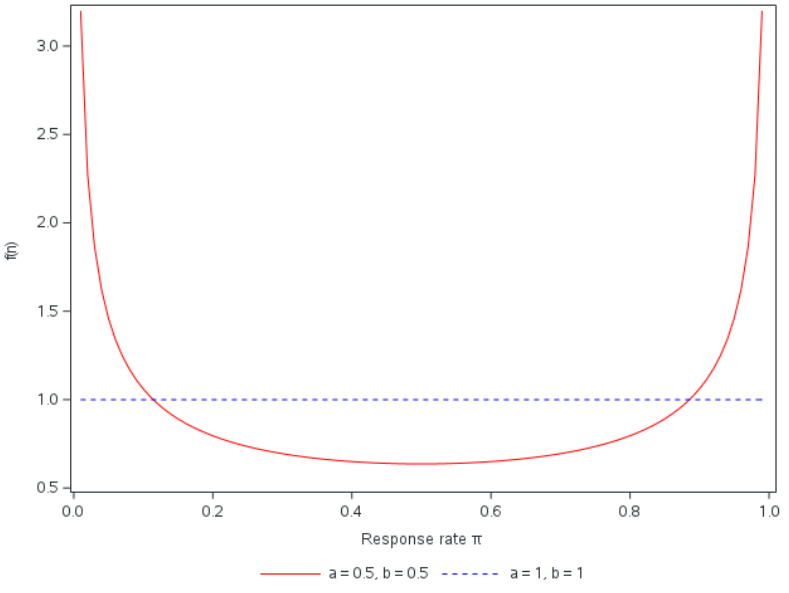

Figure 1: Beta probability density function with parameters a=0.5 and b=0.5

Source: own study.

The posterior distribution of the response rate is derived based on (1) using (2) and (4) (Lesaffre & Lawson, 2012, pp. 24-30)[3]:

$$f(\pi \mid y_1, y_2, \ldots, y_n) = \frac{L(y_1, y_2, \ldots, y_n \mid \pi) f(\pi \mid a, b)}{\int_0^1 L(y_1, y_2, \ldots, y_n \mid \pi) f(\pi \mid a, b) d\pi} = \frac{\Gamma(a+b+n)}{\Gamma(a+r)\Gamma(b+n-r)} \pi^{a+r-1}(1-\pi)^{b+n-r-1}, \tag{5}$$

which is a Beta distribution with parameters $a`=a+r$ and $b`=b+n-r$. The observed number of responses corrects the prior belief about the response rate distribution.

For example, if we observe r=10 responses in a sample of n=50 subjects, the prior beta distribution with parameters a=0.5 and b=0.5, would give the beta posterior with $a`$=10.5 and $b`$=40.5 shown in Figure 2. The empirical response rate equals π=0.2, and we can clearly see that the posterior distribution is heavily concentrated around that value.

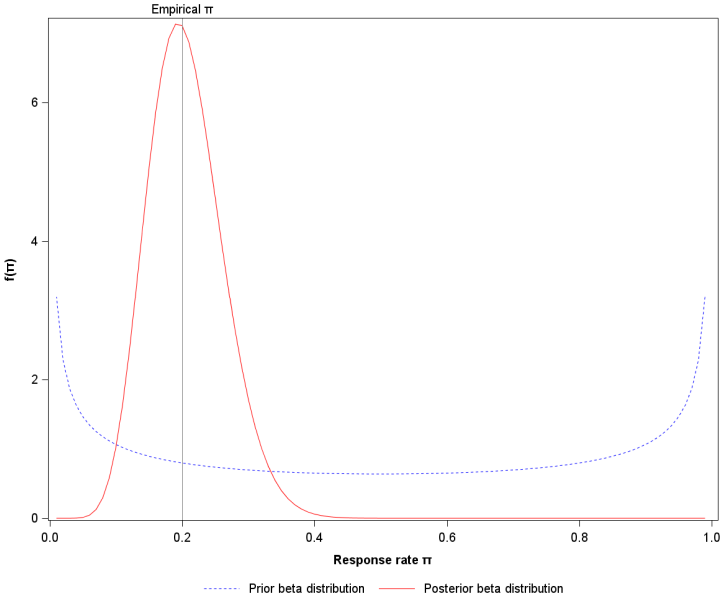

Figure 2: Beta prior distribution with a=0.5 and b=0.5 and posterior with a'=10.5 and b'=40.5

Source: own study.

The properties of Bayesian estimate of the response rate allows to make the judgement on the final outcome based on the actual number of responses at a given time point. We estimate the probability of any future outcome using posterior predictive distribution.

[3] Derivation of the posterior distribution of the response rate is included in the Appendix.

Firstly, let us note that the posterior distribution $f(\theta|\, y_1, y_2, \ldots, y_n)$ is our belief on the probability of potential values of θ given data. The posterior probability mass can be used to assess the future responses. The distribution of the future responses for continuous variable given data is (Lesaffre & Lawson, 2012, p. 53):

$$f(y_{t+1}, y_{t+2}, \ldots, y_n \mid y_1, y_2, \ldots, y_t) = \int f(y_{t+1}, y_{t+2}, \ldots, y_n \mid \theta) f(\theta \mid y_1, y_2, \ldots, y_t) d\theta. \quad (6)$$

In the settings of the response rate, the predictive posterior distribution given r_0 responses in n_0 subjects, defines the probability of observing any possible number of responses in the remaining subjects $m=n-n_0$. The number of responses in future m subjects is random binomial variable (3). Knowing that the posterior distribution of the response rate is given by (5), the predictive distribution for the response rate is:

$$\begin{aligned} P(X = x \mid r_0, n_0, a, b) &= \int_0^1 \binom{m}{x} \pi^x (1-\pi)^{m-x} \frac{\Gamma(a+b+n_0)}{\Gamma(a+r_0)\Gamma(b+n_0-r_0)} \pi^{a+r_0-1}(1-\pi)^{b+n_0-r_0-1} d\pi \\ &= \binom{m}{x} \frac{\Gamma(a+b+n_0)}{\Gamma(a+r_0)\Gamma(b+n_0-r_0)} \frac{\Gamma(a+r_0+x)\Gamma(b+n_0-r_0+m-x)}{\Gamma(a+b+n_0+m)} \quad for\ x = 0,1,\ldots, m. \end{aligned} \quad (7)$$

Let us assume that the maximum sample size in a study equals n=50. At a given timepoint we observed 21 subjects, 10 of which had a response. The current response rate equals π_0=10/21≈0.476. Figure 3 shows the probability of future responses in m remaining subjects. As we can see the mass of the probability concentrates around 14. The final response rate would then most likely be close to π = (10+14)/50=0.48.

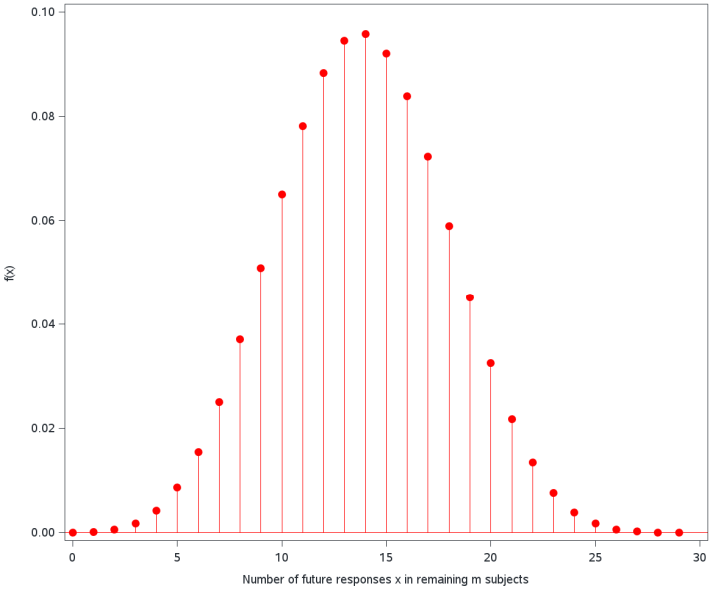

Figure 3: Posterior predictive distribution of responses in the remaining subjects with beta prior with parameters a=0.5 and b=0.5

Source: own study.

The procedure which enables to project the future results based on the current responses is built on the following steps defined in (Lee & Liu, 2008):

1. Find the posterior predictive distribution of future responses. The derivations involve the response rate π_0 computed for the number of responses r_0 and the number of subjects n_0 observed until given time point.
2. Calculate the probabilities p_x x=0,1,…,m of observing any potential number of responses x for the remaining m=n-n_0 subjects to assess how likely any possible future result is.
3. Assess the probability of observing the pre-specified response rate at each number of future responses (for example $\pi > 0.5$). This tells us how likely it is to reach the endpoint given the number of future responses. The individual probability of reaching the response rate is given by the posterior beta distribution (5) at each x.
4. Identify the number of responses for which the probability of reaching the pre-specified response rate is greater than $p_{\min}$ (e.g. $p_{\min} = 0.9$).
5. Sum the probabilities $p_x > p_{\min}$. The sum of these probabilities yields the predictive probability p of observing the pre-specified response rate when the maximum sample size is reached.

If the predictive probability is very low $p<p_L$ or very high $p>p_U$ the decision would be to stop the trial respectively for futility or efficacy. The lower bound p_L represents the probability of an event that is very unlikely whereas p_U represents the probability of an event which is very likely to be observed.

An example choice for the probability bounds is $p_L = 0.1$ and $p_U = 0.9$. In the context of the response rate assessment for $p<p_L$ the chance of reaching the requested response given the current results is very unlikely. On the contrary, when $p>p_U$ it is highly probable that the expected response rate will be reached and therefore we should consider extension of the study and moving into the large sample phase.

Continuing the example of the trial with n=50 subjects and 10 responses in 21 observed patients, we can calculate the probability of reaching the expected response rate π at the end of the trial for each number of future responses x.

Let us assume that the response is expected to occur among more than half of the patients. The probability that $P(\pi > 0.5|r_0, x)$ is calculated using the cumulative distribution function of the posterior beta distribution (5):

$$F(z \mid \mathbf{y}_0) = \begin{cases} 0 & z < 0, \\ \dfrac{\Gamma(a+b+n_0+m)}{\Gamma(a+r_0+x)\Gamma(b+n_0-r_0+m-x)} \displaystyle\int_0^z \pi^{a+r_0+x-1}(1-\pi)^{b+n_0-r_0+m-x-1} d\pi & 0 \le z \le 1, \\ 1 & z > 1. \end{cases} \quad (8)$$

where z represents the probability threshold determined by the expectation for the response rate and $\mathbf{y}_0$ reflects the characteristics of the process observed at time of the interim including r_0 and n_0.

The individual p_x along with the associated probabilities of reaching the expected response rate are shown in Table 1. Given that we observed 21 subjects the future number of responses can take a value of x=0,1,…,29, with the total number of subjects of n=50. The individual probabilities p_x are calculated based on (7) whereas probabilities $P(\pi > 0.5|r_0, x)$ are computed using (8). The table contains also the indicator function which takes the value of 1 for each number of responses for which it is highly likely that the pre-defined response threshold will be attained when the maximum sample size is reached.

Table 1. Posterior probabilities for each number of responses in m remaining subjects

x	p_x	$P(\pi > 0.5\|r_0, x)$	$I(x)$
0	1.816E-05	5.88E-06	0
1	0.00014	2.35E-05	0
2	0.0005853	8.42E-05	0
3	0.0017558	0.000271	0
4	0.0042212	0.00079	0
5	0.0086207	0.002095	0
6	0.0154922	0.005086	0
7	0.0250717	0.011339	0
8	0.0371253	0.023309	0
9	0.0508755	0.044338	0
10	0.0650539	0.078308	0
11	0.0780847	0.128848	0
12	0.088359	0.198193	0
13	0.094538	0.286031	0
14	0.0958121	0.38882	0
15	0.0920547	0.5	0
16	0.0838356	0.61118	0
17	0.0722937	0.713969	0
18	0.058906	0.801807	0
19	0.0452069	0.871152	0
20	0.0325269	0.921692	1
21	0.0218038	0.955662	1
22	0.0135001	0.976691	1
23	0.0076305	0.988661	1
24	0.0038731	0.994914	1
25	0.0017241	0.997905	1
26	0.0006494	0.99921	1
27	0.0001951	0.999729	1
28	4.181E-05	0.999916	1
29	4.826E-06	0.999976	1

Source: own study.

The probability of reaching the expected response rate is assessed given the current number of responses r_0 in n_0 subjects. In order to ensure that the goal of the study was met we would expect at least 20 responses in future patients. However, the probability

of observing these many success cases in the example setting is low. To be precise, the predicted probability of observing more than 50% of responses at the end of the trial equals $p = \sum_x I(x)p_x = 0.082$. If we took the lower bound for the predicted probability of p_L =0.1, then the conclusion would be to stop the study at the current stage for futility as $p < p_L$.

3. Empirical evidence – disease control at 12 weeks in patients with metastatic prostate cancer

The Bayesian predictive probability design has been applied to the data collected in metastatic prostate cancer clinical trial[4]. The aim of the trial was to assess the overall survival of patients with metastatic prostate cancer on standard and experimental combination therapies.

In order to exemplify the application and assess the usefulness of the Bayesian predictive probability design in clinical trial settings the following research problem has been undertaken. The purpose of the study is to find the number of patients with disease control at 12 weeks. The endpoint of the disease control rate at landmark time is based on binary outcome which simplifies the Bayesian settings. This type of endpoint finds application in phase II trials (Simon, Geyer, Subramanian & Roychowdhury, 2016, p. 18).

Disease control is defined as the number of subjects with complete response, partial response or stable disease as specified in the response criteria in the clinical study protocol for the prostate cancer clinical trial (Project Data Sphere, 2008, pp. CSP, p. 49)[5]. The disease control has been assessed at 12 weeks allowing for 2 week time window.

In our example the interim analyses have been carried out every 10 patients, until the maximum sample size has been reached. At each interim analysis the number of patients with disease control has been calculated and compared against the Bayesian predictive probability bounds, set up for the response rate at 12 weeks of 30% (π=0.3). In other words, we want to know what the likelihood of observing at least 30% of disease control patients at 12 weeks within all sampled patients is, given the actual data at time t.

The exercise has been carried out based on the n=203 patients with at least one target tumor lesion measurement. Bayesian predictive probability bounds have been computed based on the procedure outlined on page 191 using a Beta prior with parameters a=0.5 and b=0.5 for the response rate π, and the critical values for the

[4] The data are provided by CEO Roundtable on Cancer's Life Sciences Consortium, a free digital library with historical patient level data from cancer clinical trials: https://www.projectdatasphere.org/projectdatasphere/html/about, Accessed March 26, 2017.

[5] For overview of the RECIST criteria for efficacy endpoints in oncology, including disease control rate see for example: (George, Wang, & Pang, 2016, pp. 7-8).

probabilities $p_L = 0.1$ and $p_U = 0.9$. The results are presented in Table 2. In addition, the table provides the number of disease control patients along with the disease control rate and the dates expressing the recruitment process.

In the first row of Table 2 we see that the first interim disease control assessment was carried out for 10 patients in the study recruited within 11 weeks between 3rd Jan 2007 and 20th Mar 2007. Four out of ten patients reached disease control at 12 weeks, which gave disease control rate of 40% at that time point.

If we had observed only one patient with disease control at that time, then we would have had very low probability of reaching the target π=30% disease control rate in the maximum sample size. On the contrary, if we had observed six and more patients with disease control, it would have been a strong indication for efficacy.

A disease control between 2 and 5 would have positioned the investigator in the region where there is no strong indication for either of the two decisions. The indication of the model is therefore to continue the study because the current data do not provide enough statistical evidence about efficacy of the treatment.

From t=3 and n=30, which was 13 months prior to the enrolment of the last subject and until the end of the recruitment, the observed response rate was in the efficacy region. The data for 30 patients provided enough statistical evidence within the Bayesian design to conclude that the disease control rate at 12 weeks for the maximum sample size would be equal to at least 30%. The actual response rate for the maximum sample size was 90/203=44%.

Table 2. Bayesian predictive probability boundaries with actual disease control assessments at 12 weeks and recruitment time for each sample in prostate cancer clinical trial

	Recruitment			Bayesian predictive probability boundaries			Efficacy assessments	
Sample size at time t	Date first subject enrolled in each sample	Date last subject enrolled in each sample	Recruitment time in weeks	Futility region	Continuation region	Efficacy region	Number of disease control subjects	Disease control rate
10	2007-01-03	2007-03-20	11	1	2-5	6-10	**4**	40%
20	2007-03-26	2007-05-25	8	1-4	5-9	10-20	**9**	45%
30	2007-05-31	**2007-06-26**	4	1-7	8-13	14-30	**15**	50%
40	2007-07-02	2007-07-26	3	1-10	11-17	18-40	**18**	45%
50	2007-07-30	2007-09-03	5	1-13	14-20	21-50	**25**	50%
60	2007-09-05	2007-09-28	3	1-16	17-24	25-60	**29**	48%
70	2007-10-01	2007-10-17	2	1-19	20-28	29-70	**31**	44%
80	2007-10-23	2007-11-20	4	1-23	24-31	32-80	**34**	43%
90	2007-11-22	2007-12-13	3	1-26	27-35	36-90	**37**	41%
100	2007-12-14	2008-01-07	4	1-29	30-38	39-100	**44**	44%
110	2008-01-08	2008-02-08	4	1-33	34-41	42-110	**47**	43%

Table 3: Bayesian predictive probability boundaries with actual disease control assessments at 12 weeks and recruitment time for each sample in prostate cancer clinical trial (cont.)

Sample size at time *t*	Recruitment			Bayesian predictive probability boundaries			Efficacy assessments	
	Date first subject enrolled in each sample	Date last subject enrolled in each sample	Recruitment time in weeks	Futility region	Continuation region	Efficacy region	Number of disease control subjects	Disease control rate
120	2008-02-14	2008-03-11	4	1-36	37-45	46-120	**53**	44%
130	2008-03-14	2008-03-28	2	1-40	41-48	49-130	**56**	43%
140	2008-04-04	2008-04-21	3	1-43	44-51	52-140	**61**	44%
150	2008-04-24	2008-05-19	4	1-47	48-55	56-150	**66**	44%
160	2008-05-20	2008-06-11	3	1-51	52-58	59-160	**69**	43%
170	2008-06-11	2008-07-07	4	1-54	55-61	62-170	**76**	45%
180	2008-07-08	2008-08-04	4	1-58	59-64	65-180	**79**	44%
190	2008-08-06	2008-09-15	6	1-62	63-67	68-190	**83**	44%
200	2008-09-16	2008-10-08	3	1-67	68-69	70-200	**88**	44%
203	2008-10-10	**2008-10-17**	1				**90**	44%

Source: own study based on (Project Data Sphere, 2008).

The number of subjects with disease control at the subsequent interim assessments from Table 2 has been illustrated in Figure 4.

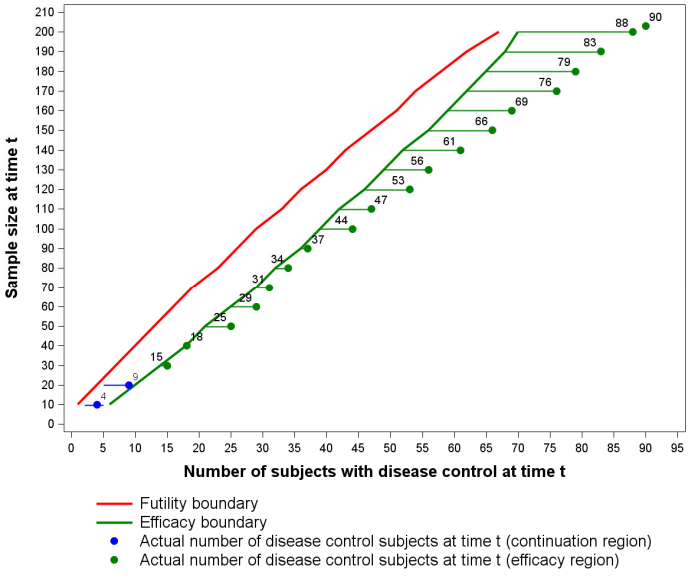

Figure 4: Disease control rate at week 12 in patients with metastatic prostate cancer and the Bayesian predictive boundaries assessed at every 10 patients in the study

Source: own study based on Table 2.

4. Sensitivity analysis – the choice of prior distribution

As noted earlier the Bayesian analysis allows bringing an external expertise into the estimation process. In our example we could look for evidence in the form of response rates from previous studies in metastatic prostate cancer. This evidence would serve as input for defining the parameters of the prior distribution.

In the example in Section 2 the target response rate is at least π=0.3 of disease control patients. In order to assess the sensitivity of the Bayesian design to the selection of prior distribution two extreme scenarios have been applied, additionally to the non-informative prior from the above example.

Figure 5 shows the posterior distribution resulting from prior assumption of poor performance of the treatment (red solid line for a=1 and b=10). As compared to the non-informative scenario (orange solid line for a=0.5 and b=0.5) the pessimistic posterior is shifted to the left, towards the lower response rate. We would then require a lot more evidence in the form of response from the new trial in order to claim efficacy (see Table 4, first section '*Low prior response rate*'). What is more, we would quicker consider the therapy futile, e.g. 3 responses out of 10 would result in futility decision. The opposite can be observed for the reverse selection of the prior (see Figure 6), however it must be noted that this scenario is far more extreme than the previous pessimistic one, and was applied only to provide a full overview over the resulting trial designs.

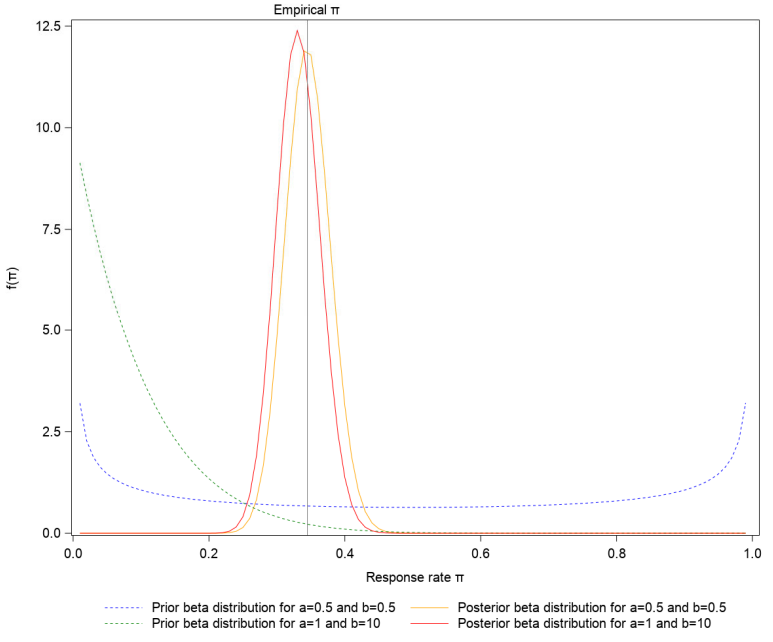

Figure 5: Beta prior distributions with a=0.5 and b=0.5, and a=1 and b=10 with resulting posterior distributions and assumption of empirical response rate of π=70/203≈0.34

Source: own study.

Figure 6 shows the posterior distribution resulting from prior assumption of very good performance of the treatment (red solid line). As noted this is an extreme case in the sense that we kept the target response rate the same at the level of approximately 30%. However, the prior would suggest far higher response rates in the former trials.

The resulting design would drive the efficacy lower boundary to the minimum around slightly above 30% (see Table 4, third column '*High prior response rate*'). We would relatively quickly consider the trial efficacious, e.g. for 3 responses out of 10 first patients.

Considering the above cases, the choice of the prior distribution is highly affecting the decision about the study. This observation is very clear for early stages of the study conduct and for low number of patients. Given the range of possible results the selection of informative prior for low number of patients would require very strong argument. On the other hand, the sensitivity to the choice of the prior distribution is to some extent diminishing for larger number of subjects, i.e. for later stages of the trial. For example, at the time of having 160 patients the lower efficacy bound is 62/160≈0.39 for conservative prior and 54/160≈0.34 for optimistic prior.

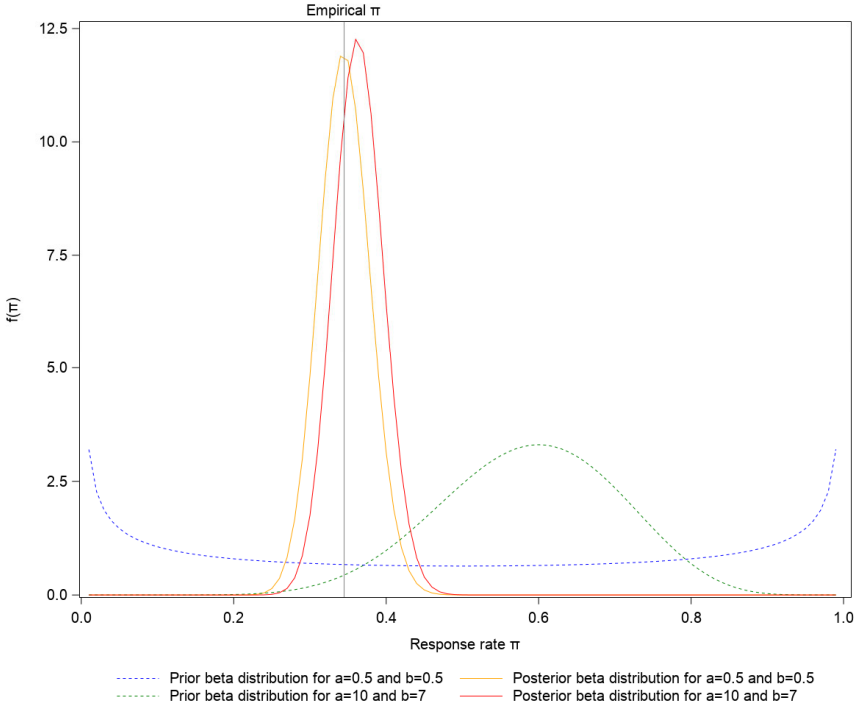

Figure 6: Beta prior distributions with a=0.5 and b=0.5, and a=10 and b=7 with resulting posterior distributions and assumption of empirical response rate of $\pi=70/203\approx0.34$

Source: own study.

Table 4: Bayesian predictive probability boundaries designed for prostate cancer clinical trial according to three types of prior distribution

Sample size at time *t*	Bayesian predictive probability boundaries ***Low prior response rate*** Beta prior with *a*=1 and *b*=10			Bayesian predictive probability boundaries ***Non-informative prior*** Beta prior with *a*=0.5 and *b*=0.5			Bayesian predictive probability boundaries ***High prior response rate*** Beta prior with *a*=10 and *b*=7		
	Futility region	Continuation region	Efficacy region	Futility region	Continuation region	Efficacy region	Futility region	Continuation region	Efficacy region
10	1-3	4-8	9-10	1	2-5	6-10		1-2	3-10
20	1-6	7-12	13-20	1-4	5-9	10-20		1-6	7-20
30	1-9	10-16	17-30	1-7	8-13	14-30	1-2	3-9	10-30
40	1-12	13-20	21-40	1-10	11-17	18-40	1-5	6-13	14-40
50	1-16	17-24	25-50	1-13	14-20	21-50	1-8	9-16	17-50
60	1-19	20-27	28-60	1-16	17-24	25-60	1-11	12-20	21-60
70	1-22	23-31	32-70	1-19	20-28	29-70	1-15	16-23	24-70
80	1-25	26-34	35-80	1-23	24-31	32-80	1-18	19-27	28-80
90	1-29	30-38	39-90	1-26	27-35	36-90	1-21	22-30	31-90
100	1-32	33-41	42-100	1-29	30-38	39-100	1-25	26-34	35-100
110	1-36	37-45	46-110	1-33	34-41	42-110	1-28	29-37	38-110
120	1-39	40-48	49-120	1-36	37-45	46-120	1-32	33-40	41-120
130	1-43	44-51	52-130	1-40	41-48	49-130	1-35	36-44	45-130
140	1-46	47-54	55-140	1-43	44-51	52-140	1-39	40-47	48-140
150	1-50	51-58	59-150	1-47	48-55	56-150	1-42	43-50	51-150
160	1-54	55-61	62-160	1-51	52-58	59-160	1-46	47-53	54-160
170	1-57	58-64	65-170	1-54	55-61	62-170	1-50	51-56	57-170
180	1-61	62-67	68-180	1-58	59-64	65-180	1-53	54-59	60-180
190	1-65	66-70	71-190	1-62	63-67	68-190	1-57	58-62	63-190
200	1-70	71-72	73-200	1-67	68-69	70-200	1-62	63-64	65-200
203									

Source: own study.

An example of the analysis incorporating the prior knowledge from findings in previous trials is provided in (Heath, et al., 2020; Chen, et al., 2019).

5. Summary

The theoretical properties of the Bayesian predictive probability design are appealing as a tool for detecting the treatment signal at earlier stages of studies with continuous recruitment of subjects to the sample. The practical application has shown the usefulness of the approach from the perspective of the timing of the decision. This goes along with the known argument for adaptive design allowing for reducing the overall sample size, cost of the study, drug development time length (George, Wang, & Pang, 2016, p. 367). The final decision would still require larger programs in terms of the sample size. It that sense the Bayesian design has a supportive role.

The results are affected by the level of the expected response rate and therefore the choice of that parameter is crucial for the analysis and conclusions. When discussing the application of Bayesian predictive design, the expected time to response on specific endpoint must also be considered (Zhou, Liu, Kim, Herbst, & Lee, 2008). The expectation is to use a conservative level of the response rate and refer to the findings in other clinical studies with similar indication in order to formulate the response rate that would be clinically beneficial. The other question arose in the analysis is how to define the appropriate frequency of the assessments in order to draw meaningful conclusions in possibly short time period.

From statistical perspective the interest is also in the choice of prior distribution, which affects the expected response rate and, as a consequence, the interim decision. The presented example showed the sensitivity of the prior assumption to the resulting predictive design. A very careful considerations would be required for the choice of informative prior distribution, especially in view of decision making at early stages of the recruitment in the trial. From broader perspective, the prior assumptions are vital for the trial assessment and it is important to consider how robust the design is in translating into phase III trial (Harrington & Parmigiani, 2016, p. 8).

Moreover, there are many practical problems with predictive design related to the conduct of a clinical trial and data collection associated with that process. Firstly, we need to decide on the subjects to be reviewed at each stage. Due to practical problems related to data collection the enrolment date might not be enough to decide about the assignments of the subjects to be included in the interim analysis. There would normally be operational reasons for specifying the enrolment stage at which the interim is planned (Heath, et al., 2020, p. 3). The practical aspects would have to be taken into account within the schedule determined through statistical considerations.

One can argue that by setting the sample size for early review too low could lead to a number of trials inappropriately stopped (Mitchell, 2018, p. 300). Practically, with a small number of patients the investigators would refrain from entering a large sample phase even with very promising results.

Lastly, the Bayesian predictive design requires specification of the criteria for inclusion in the interim analysis. Especially in cancer clinical trials we may expect deviations from the protocol-defined procedures which can bias the final result and therefore need to be carefully considered in the analysis.

References

Lee, J., Liu, D., (2008, 5). A predictive probability design for phase II cancer clinical trials. *Clinical Trials*, pp. 93–106.

Little, J. A., Rubin, D., (2002). *Statistical Analysis with Missing Data.* Hoboken: John Wiley & Sons.

Mittelhammer, R., (2013). *Mathematical Statistics for Economics and Business.* New York: Springer.

Lesaffre, E., Lawson, A., (2012). *Bayesian Biostatistics.* Chichester: John Wiley & Sons.

Jóźwiak, J., Podgórski, J., (2009). *Statystyka od podstaw.* Warszawa: Polskie Wydawnictwo Ekonomiczne.

Bolstad, W., (2007). *Introduction to Bayesian Statistics.* Hoboken: John Wiley & Sons.

Project Data Sphere, (2008, July 21). *A Randomized, Open Label Multi-Center Study of XRP6258 at 25 mg/m^2 in Combination With Prednisone Every 3 Weeks Compared to Mitoxantrone in Combination With Prednisone For The Treatment of Hormone Refractory Metastatic Prostate Cancer*. Retrieved March 7, 2017, from https://www.projectdatasphere.org/projectdatasphere/html/content/79.

George, S., Wang, X., & Pang, H., (2016). *Cancer Clinical Trials. Current and Controversial Issues in Design and Analysis.* Boca Raton: CRC Press.

Heath, A., Offringa, M., Pechlivanoglou, P., J. D., R., Klassen, T., Poonai, N., & Pullenayegum, E., (2020). Determining a Bayesian predictive power stopping rule for futility in a non-inferiority trial with binary outcomes. *Contemporary Clinical Trials Communications, 18*.

Zhou, X., Liu, S., Kim, E., Herbst, R., & Lee, J., (2008). Bayesian adaptive design for targeted therapy development in lung cancer--a step toward personalized medicine. *Clinical Trials, 5*(3), pp. 181–193.

Mitchell, P., (2018). A Bayesian single-arm design using predictive probability monitoring. *Biom Biostat Int J., 7*(4), pp. 299–309.

Barker, A., Sigman, C., Kelloff, G., Hylton, N., Berry, D., & Esserman, L., (2009). I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther., 86*(1), pp. 97–100.

Simon, R., Geyer, S., Subramanian, J., & Roychowdhury, S., (2016). The Bayesian basket design for genomic variant-driven phase II trials. *Semin Oncol., 43*(1), pp. 13–18.

Chen, M., Ibrahim, J. G., Lam, P., Yu, A., & Zhang, Y., (2011). Bayesian Design of Noninferiority Trials for Medical Devices Using Historical Data. *Biometrics, 67*(3), pp. 1163–1170.

Yin, F., Chen, N., & Lee, J., (2012). Phase II trial design with Bayesian adaptive randomization and predictive probability. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 61*(2), pp. 219–235.

Harrington, D., Parmigiani, G., (2016). I-SPY 2 — A Glimpse of the Future of Phase 2 Drug Development? *The New England Journal of Medicine, 375*(1), pp. 7–9.

Chen, D., Schell, M., W. J., F., Pettersson, F., Kim, S., J. E., G., & E.B., H., (2019). Application of Bayesian predictive probability for interim futility analysis in single-arm phase II trial. *Transl Cancer Res., 8*, pp. 404–420.

Appendix A

Posterior distribution of the response rate with Beta distribution as prior:

$$
\begin{aligned}
f(\pi \mid y_1, y_2, \ldots, y_n) &= \frac{L(y_1, y_2, \ldots, y_n \mid \pi) f(\pi \mid a, b)}{\int_0^1 L(y_1, y_2, \ldots, y_n \mid \pi) f(\pi \mid a, b) d\pi} \\
&= \frac{\pi^r (1-\pi)^{n-r} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1}}{\int_0^1 \pi^r (1-\pi)^{n-r} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1} d\pi} \\
&= \frac{\pi^{a+r-1} (1-\pi)^{b+n-r-1}}{\int_0^1 \pi^{a+r-1} (1-\pi)^{b+n-r-1} d\pi} \\
&= \left(\frac{\Gamma(a+r)\Gamma(b+n-r)}{\Gamma(a+b+n)} \right)^{-1} \pi^{a+r-1} (1-\pi)^{b+n-r-1} \\
&= \frac{\Gamma(a+b+n)}{\Gamma(a+r)\Gamma(b+n-r)} \pi^{a+r-1} (1-\pi)^{b+n-r-1}.
\end{aligned}
\tag{9}
$$