Bayesian Hypothesis Testing (BHT) in Single-arm Phase II Clinical Trials with Binary Endpoints

Martin Kappler[1], Joachim Gerß[2]

[1]Novartis Pharma S.A.S., France
[2]Institute of Biostatistics and Clinical Research, University of Münster, Germany

martin.kappler@novartis.com

Introduction

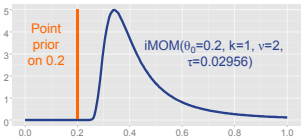
When evaluating the response rate p of a new therapeutic agent using a binary endpoint in a single-arm phase II clinical trial, two competing hypotheses are established, the null hypothesis H_0: $p \le p_0$ and the alternative hypothesis H_1: $p > p_0$. The final aim of the statistical analysis is to draw a decision either in favor of efficacy of the trial agent (rejecting H_0) or futility. The problem is usually solved in a sequential approach. Stopping boundaries may be defined using frequentist methods, e.g. Simon's two-stage design [1]. In a Bayesian framework different approaches may be pursued. Stopping boundaries may be derived from posterior and predictive probabilities (Bayesian estimation). Recently, Johnson and Cook proposed an alternative approach based on formal hypothesis testing (Bayesian hypothesis testing, BHT) [2]. In the present work based on the BHT approach a new and – as will be shown – more efficient algorithm is derived. All above approaches are evaluated and compared with respect to their operating characteristics using simulated data.

Methods

Consider a single-arm clinical trial with binary endpoint. The futility level of the response rate p is p_0=0.2 and the target rate is p_1=0.4. The following hypotheses are tested H_0: $p \le p_0$ and H_1: $p > p_0$ using a one-sided type I error of 5%. At least 80% power is wanted under the target rate p_1.

Bayesian hypothesis testing (BHT)

Inverse moment (iMOM) prior under H_1: In a recent approach proposed by Johnson and Cook [2] different prior distributions for the response rate p are assigned under the null and alternative hypothesis, respectively. A one-point prior density is used under H_0. The prior density under H_1 is chosen to be of the class of *inverse moment densities* (iMOM). No probability mass is assigned to parameter values that are consistent with H_0 (nonlocal alternative prior density). The density is parameterized, so that the median of the iMOM density is p_1=0.4. The authors claim that the approach has the useful property that mis-specification of the prior density (e.g. over-optimistic) cannot increase the expected weight of evidence in favor of the trial agent.



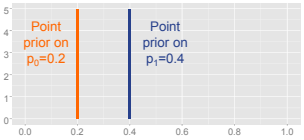
Data model: $X_n|p \sim Bin(n,p)$
Prior distribution under H_0 (**orange**) and under H_1 (**blue**, iMOM prior):
Hyperprior: $Prob(H_0) = Prob(H_1) = 0.5$

Patient outcomes are monitored continuously from a pre-determined minimal sample size N_{min}=10 up to a final sample size N_{max}=50. Patients are accrued successively and after each observed outcome the posterior probabilities of the hypothesis H_0 and H_1 are calculated. The trial can be stopped either for efficacy or futility.

- Stop for efficacy and rejection of H_0: The posterior probability $Prob(H_1 \mid x_n,n)$ is greater than a cutoff c_{eff}.
- Stop for futility and no rejection of H_0: The posterior probability $Prob(H_0 \mid x_n,n)$ is greater than a cutoff c_{fut}.

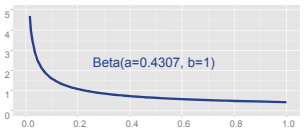
The cutoffs are fixed at c_{eff} = 0.9 and c_{fut} = 0.9. If in the final analysis neither $Prob(H_0 \mid x_{Nmax},N_{max}) > c_{fut}$ nor $Prob(H_1 \mid x_{Nmax},N_{max}) > c_{eff}$, no decision is drawn.

Point prior under H_1: In the present work a new algorithm is derived based on the above BHT approach using a one-point prior density not only under H_0 (**orange**) but also under H_1 (**blue**). The cutoffs for efficacy and futility stop are fixed at c_{eff} = 0.925 and c_{fut} = 0.9, respectively.



Bayesian estimation

The BHT approach is compared to a Bayesian estimation approach based on the Beta-Binomial model. The decision for efficacy or futility is not based on the probabilities of the hypotheses H_0 and H_1 but on the posterior distribution (efficacy) and on the predictive distribution for the final response rate (futility).



Data model: Number of responders among n patients $X_n|p \sim Bin(n,p)$
Minimally informative unimodal Beta prior (**blue**) for p with median p_0 = 0.2:

As in the BHT approach, patient outcomes are monitored continuously from a minimal sample size N_{min}=10 up to a final sample size N_{max}=50. After each successively observed outcome the number of responses x_n is used to update the posterior distribution of the response rate (beta distribution) and the predictive distribution of the number of responders among N_{max} patients (beta-binomial distribution). The trial can be stopped either for efficacy or futility.

- Stop for efficacy and rejection of H_0: a) The posterior probability $Prob(p>p_0|x_n,n)$ is greater than a cutoff c_{eff} AND b) the observed response rate x_n/n is greater than a minimally required observed rate. The cutoff c_{eff} is fixed at 0.9875 and the minimally required observed rate is chosen as the mean between p_0 and p_1 (i.e. $(p_0+p_1)/2$ = 0.3).
- Stop for futility and no rejection of H_0: a) The predictive probability $Prob(x_{Nmax}/N_{max} \le p_0 \mid x_n,n)$ (final observed response rate being lower than the futility level p_0) is greater than a cutoff c_{fut1} OR b) The predictive probability $Prob(x_{Nmax}/N_{max} \ge (p_0+p_1)/2 \mid x_n,n)$ (final observed response rate being higher than the minimally required observed rate) is lower than a cutoff c_{fut2}. The cutoffs are fixed at c_{fut1} = 0.9 and c_{fut2} = 0.1.

Simon's two-stage design [1]

As a reference for the above approaches, Simon's two-stage design is used. This approach is based on frequentist test theory and exact binomial probabilities. One interim analysis with n patients is scheduled with an optional futility stop if the number of responders x_n is too small, otherwise recruitment continues up to the pre-determined final sample size N.

Optimal design: Interim analysis: Futility stop if $x_n \le 3$, n = 13; Final analysis: Decision for efficacy if $x_N \ge 13$, N = 43

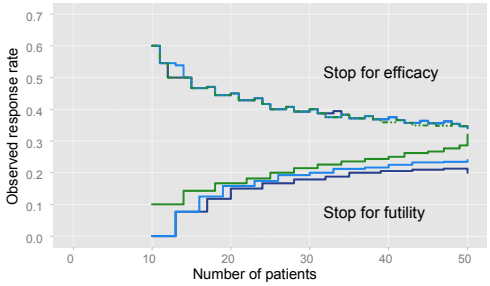
Minimax design: Interim analysis: Futility stop if $x_n \le 4$, n = 18; Final analysis: Decision for efficacy if $x_N \ge 11$, N = 33

Comparison of approaches

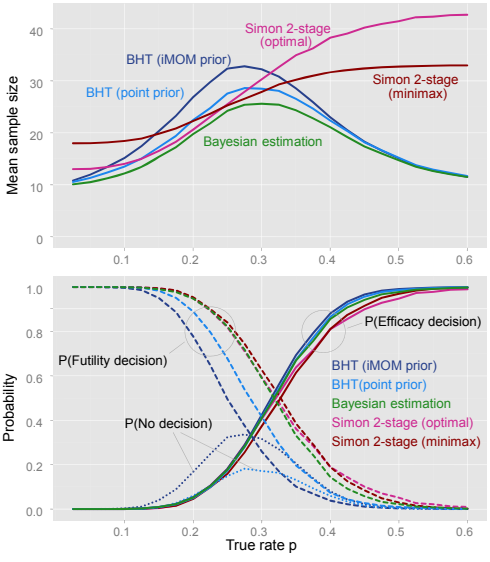
In order to perform a fair comparison between all above approaches a standardization was necessary. Therefore frequentist criteria were adopted. The different cutoffs were chosen to match a type I error of 5% and a maximum type II error of 20% for the targeted rate p_1=0.4. In spite of this standardization it has to be noted that Simon's design fundamentally differs from the proposed Bayesian designs. Only one interim analysis is scheduled instead of continuous monitoring and no early stopping for efficacy is possible. Both aspects might be introduced in modified versions of Simon's design. However, in the present work the original designs were considered as reference due to their widespread use.

Results

The stopping boundaries of all Bayesian approaches are quite similar with regard to the upper boundary (stop for efficacy) but they show differences on the lower boundary (stop for futility).



The operating characteristics of all approaches are compared using simulated data (5000 trials for each rate). Presented are the mean sample size, the probability of deciding in favor of efficacy of the trial agent (rejecting H_0) or futility, as well as the probability of drawing no decision.



Conclusion

As expected all Bayesian approaches differ markedly from the Simon's two-stage designs. However, it has to be noted that Simon's designs fundamentally differ from the proposed Bayesian designs and serve only as reference.

In the BHT approach, using a one-point prior density under H_1 is more efficient than using an iMOM prior density. The mean sample size is lower and for low response rates p a decision in favor of futility is drawn more often instead of drawing no decision. However, for rates between p_0 and p_1 high probabilities for no decision are preferable and the iMOM prior shows better behavior.

Bayesian estimation shows better mean sample size and a higher probability of deciding in favor of futility if H_0 is true compared to both BHT approaches.

Overall, the proposed Bayesian estimation approach seems to be a good alternative to the BHT approaches. As further advantage of this approach, point and interval estimates can be easily obtained based on the posterior distribution of p.

Literature

1 Simon R (1989). Optimal two-stage designs for phase II clinical trials. Controlled Clinical Trials 10, 1-10.

2 Johnson VE, Cook JD (2009). Bayesian design of single-arm phase II clinical trials with continuous monitoring. Clinical Trials 6, 217-6.