A Predictive Probability Design for Phase II Cancer Clinical Trials

S-PLUS/R Program Read Me File

This document describes the computational issues in the manuscript "A predictive probability design for phase II cancer clinical trials" by J. Jack Lee and Diane Liu

Under the hypothesis testing framework, a phase IIA clinical trial is designed to test

H_0: $p \le p_0$

H_1: $p \ge p_1$

where p_0 represents a pre-specified response rate given the standard treatment and p_1 represents a target response rate to a new treatment. A study is designed such that

$Prob(Accept\ New\ Treatment \mid H_0) = \alpha$

and $Prob(Reject\ New\ Treatment \mid H_1) = \beta$

where α and β are type I and type II error rates, respectively. Given p_0, p_1, the maximum number of patients, number of stages, cohort size of each cohort at each stage, acceptance region and rejection region for each cohort, the type I and type II error rates, the probability of early termination (*PET*) of the trial and the expected sample size (*E(N)*) under H_0 can be calculated by applying the recursive formulas of Schultz et al. (1973).

Predictive probability (PP) Approach in A Bayesian Setting

In the Bayesian approach, we assume that the prior distribution of the response rate $\pi(p)$ follows a beta distribution, $beta(a_0, b_0)$. It represents the investigator's previous knowledge or belief of the efficacy of the new regimen. The quantity $a_0 / (a_0 + b_0)$ reflects the prior mean while size of $a_0 + b_0$ indicates how informative the prior is. The larger the value of $a_0 + b_0$, the more informative the prior and the stronger the belief it contains. We set a maximum accrual of patients to N_{max}. We assume the number of observed responses in the current n ($n \le N_{max}$) patients, X, follows a binomial distribution, $binomial(n, p)$, and the likelihood function for the observed data x is

$L_x(p) \propto p^x \times (1-p)^{n-x}$

Consequently, the posterior distribution of the response rate follows a beta distribution

$$P|x \sim beta(a_0 + x,\ b_0 + n - x)$$

Thus, the number of responses in the potential $m=N_{max}-n$ future patients, Y, follows a beta-binomial distribution, $beta\text{-}binomial(m, a_0 + x, b_0 + n - x)$.

When $Y=i$, we denote the posterior probability of P as $f(p|x, Y=i)$, where

$$P|x, Y=i \sim beta(a_0 + x + i,\ b_0 + N_{max} - x - i)$$

To calculate the predictive probability, we further define

$$B_i = Prob(P > p_0 \mid x)$$

which measures the probability that the response rate is larger than p_0 given x responses in n patients in the current data and i responses in m patients in the future. Comparing B_i to a threshold value θ_T yields an indicator I_i for considering that the treatment is efficacious at the end of the trial given the current data and the potential outcome of $Y=i$. We define

$$\text{Predictive Probability (PP)} = \sum_{i=0}^{m} \{Prob(Y=i \mid x) \times (Prob(P > p_0 \mid x, Y=i) > \theta_T)\}$$

$$= \sum_{i=0}^{m} \{Prob(Y=i \mid x) \times I(B_i > \theta_T)\}$$

$$= \sum_{i=0}^{m} \{Prob(Y=i \mid x) \times I_i\}$$

where $Prob[Y=i \mid x]$ is the probability of observing i responses in future m patients given current data x. The weighted sum of indicator I_i over Y yields the predictive probability (PP) of concluding a positive result by the end of the trial based on the cumulative information in the current stage. A high PP means that the treatment is likely to be efficacious by the end of the study given the current data, whereas a low PP suggests that the treatment may not have sufficient activity. Therefore, PP can be used to determine whether the trial should be stopped early due to efficacy/futility or continued because the current data are not yet conclusive. The decision rules can be constructed as follows:

If $PP < \theta_L$, then stop the trial and reject the alternative hypothesis;

If $PP > \theta_U$, then stop the trial and reject the null hypothesis;

Otherwise continue to the next stage until reaching N_{max} patients.

Typically, we choose θ_L as a small positive number and θ_U as a large positive constant, both between 0 and 1 (inclusive). $PP < \theta_L$ indicates that it is unlikely the response rate will be larger than p_0 at the end of the trial given the current information. When this happens, we may as well stop the trial and reject the alternative hypothesis at that point. On the other hand, when $PP > \theta_U$, the current data suggest that if the same trend continues, we will have a high probability to conclude that the treatment is efficacious at the end of the study. This result, then, provides evidence to stop the trial early due to efficacy. By choosing $\theta_L>0$ and $\theta_U<1.0$, the trial can terminate early due to either futility or efficacy. For phase IIA trials, we prefer to choose $\theta_L>0$ and $\theta_U=1.0$, to allow early stopping due to futility, but not due to efficacy.

S-PLUS/R Functions

1. *p.beta.binomial()* calculates probability of observing number of responses in patients if the number of responses follows a beta-binomial distribution.

   Note that the number of responses in n patients follows a binomial distribution *binomial(n, p).* For example, if the prior of response rate p follows a beta distribution *beta(a_0, b_0);* we observed x responses in n patients, and the total planned accrual is N_{max}, then the number of responses in future N_{max}-n patients follows a beta-binomial distribution, *beta-binomial(N_{max}-n, $a_0 + x$, $b_0 + n - x$).* The probability of observing y responses in next N_{max}-n patients can be calculated as

   p.beta.binomial(y, N_{max}-n, a_0+x, b_0+n-x).

2. *predictive.prob()* calculates predictive probability of observing the response rate larger than a certain level if the study continues to the end, given the observed data. This function calls *p.beta.binomial().*

   For example, if n patients have been treated and are evaluable for response; the prior of response rate follows *beta(a_0, b_0)*; a total of N patients are accrued, then, the

predictive probability (*PP*) of declaring the response rate larger then p_0 given each of *0:n* possible responses can be calculated as:

$$predictive.prob(N.obs=n,\ N.max=N_{max},\ p.star = p_0,\ prior = c(a_0, b_0),$$
$$pL = \theta_L,\ pU = \theta_U,\ pT = \theta_T)$$

The "*decision"* column in the output object is calculated as $(PP>=\theta_L) + (PP>=\theta_U)$, with possible outcomes of 0, 1, 2, representing $PP<\theta_L$, $\theta_L <= PP < \theta_U$ and $PP>=\theta_U$, respectively. It is used for trial boundary decision making in function *trial.bound()*.

3. *trial.bound()* computes the predictive probability stopping boundaries for the pre-specified cohort sizes, using the result from *predictive.prob()*.

For example, if the cohort sizes are $n=c(n_1, n_2, n_3, n_4, n_5)$, where n_1 is the first cohort of patients being evaluated for response before the *PP* interim decision starts to be implemented, and n_5 is the same as maximum sample size N_{max}, the response rate to the standard treatment is p_0; prior distribution of response rate follows a beta distribution $beta(a_0, b_0)$. Given θ_T, θ_L, and θ_U, the stopping boundaries can be determined by

$$trial.bound(n= c(n_1, n_2, n_3, n_4, n_5),\ N=N_{max},\ p.star = p_0,\ prior = c(a_0, b_0),$$
$$pL = \theta_L,\ pU = \theta_U,\ pT = \theta_T)$$

In the output object, column *n* represents the cohort sizes of each interim decision point, columns *lobnd* and *upbnd* are the lower and higher boundaries of continuing the trial, respectively. If number of responses in *n* patients < *lobnd*, we stop the trial declaring the treatment not efficacious. On the other hand, if number of responses in *n* patients > *upbnd*, we stop the trial declaring the treatment promising.

We choose θ_L as a small positive number and θ_T, θ_U as large positive constants, between *0* and *1*. We let $\theta_L>0$ and $\theta_U<1.0$ to allow early stopping due to futility and efficacy. However, $\theta_L>0$ and $\theta_U=1.0$ will allow early termination of the trial due to futility, but not due to efficacy.

4. The output object from *cgmp()* is used in *exact.power()* to compute the probability of continuing from each stage under the boundary constraints input. This function implements the recursive formulas of Shultz et al. (1973).

5. *exact.power()* provides operating characteristics given a boundary calculated by *trial.bound()* and the true response rate. This function calls *cgmp()*. It computes the probability of early termination due to either futility or efficacy (columns *negative* and *positive* in *$p.table*, respectively) and the probability of continuing the trial at each stage. It also provides the probability of positive result at the end of the trial (*$final.pb$positive*), mean sample size (*$mean.sample.size*), and overall probability of early termination (*$PET*). If the true response rate is p_0, then *$final.pb$positive* represents type I error rate. On the other hand, if the true response rate is p_1, then *$final.pb$positive* represents the power in declaring efficacy when the true response rate is p_1.

   For example, if a boundary *bnd* is the result from *trial.bound()*, the operating characteristics under a true response rate of *p* can be computed by evaluating

   exact.power(bnd, p.true=p)

6. *search.pLpT()* searches for ranges of θ_L and θ_T that yield the highest power at a maximum tolerable type I error rate. It performs an exhaustive search and the function can take a long time to run.

   For example, we provide the following parameters:

   cohort sizes $n=c(n_1, n_2, n_3, n_4, n_5)$, where n_1 is the number of patients first being evaluated for response before the *PP* interim decision starts to be implemented, and n_5 is the same as maximum sample size N_{max},

   maximum sample size N_{max},

   the response rate to the standard treatment p_0,

   expected increase in response rate *delta*,

prior distribution of p_0, *beta*(a_0, b_0),

θ_U as *pU=1*,

the ranges of θ_T and θ_L as *pT.range* and *pL.range*, respectively,

the nominal level of type I error rate $alpha_0$, and power $power_0$

Combinations of θ_T and θ_L that yield the highest power at tolerable type I error rate can be found by

search.pLpT(n= c(n_1, n_2, n_3, n_4, n_5), N=N_{max}, p.star = p_0, prior = c(a_0, b_0), delta, pU=1, pL.range, pT.range, $alpha_0$, $power_0$, figures=T)

Option *figures=T* allows the function to generates perspective plots showing the type I error rate, power, and the expected sample size under H_0 given the combination of θ_T and θ_L. This features can be suppressed by *figures=F*.

The following parameters and results are included in the output object, if a design with controlled type I and II error rates is found: N_{max}, *p.star=p0, p1 = p0+delta, prior, pL.range, pT.range, pL.chosen, pT.chosen, boundary.chosen, power.chosen, EN1*. *pL.chosen, pT.chosen, boundary.chosen, power.chosen* and *EN1* are the chosen ranges of θ_T and θ_L that yield the highest power at tolerable type I error rate, boundaries (including operating characteristics under H_0), power and expected sample size under H_1, respectively.

7. When designing a trial, with p_0, p_1, *prior*, *n*, θ_U and nominal level of type I error rate and power fixed, the searching algorithm finds ranges of θ_L and θ_T under an N_{max} that generates a boundary with the highest power under allowed type I error rate. We search over a range of total sample size N_{max}. For example, if the total sample size from Simon's minimax two-stage design is N_s, we choose a sufficiently large range, $N_s \pm 0.20 \times N_s$ for example, so that the smallest N_{max} which controls both type I and type II error rates can be identified.

Reference:

1. *Schultz JR, Nichol FR, Elfring GL, Weed SD: Multiple-stage procedures for drug screening. Biometrics 29:293-300, 1973.*

Acknowledgement

The authors would like to thank Dr. His-Guang Sung for his input in the discussion and codes he shared with us in the early stage of this project.