Optimal selection of a futility stopping rule

MedianaDesigner package

1. Introduction

This document provides a description of the statistical methodology used in the futility module (FutRule function).

For more information on the MedianaDesigner package, visit the following web pages at

http://www.mediana.us/medianadesigner

http://medianasoft.github.io/MedianaDesigner

2. Optimal selection of a futility stopping rule

2.1. Trial design

Consider a multi-arm Phase II or Phase III trial. The trial will be conducted to investigate the efficacy and safety of several doses or regimens of an experimental treatment versus placebo. The primary efficacy endpoint is a continuous, binary or time-to-event endpoint.

A single unblinded interim analysis will be conducted in this trial. A futility stopping rule will be applied at the interim analysis to perform a futility assessment for each treatment arm versus placebo. A treatment arm will be dropped for futility if this corresponding treatment is unlikely to be effective, e.g., the predicted probability of success at the final analysis is low. A futility stopping rule is defined in Section 2.2 and an optimal futility stopping rule maximizing the rule's sensitivity and specificity rates is derived in Section 2.3. The proposed approach to defining optimal futility stopping rules is illustrated in Section 3.

2.2. Futility stopping rule

The futility stopping rule at the interim analysis could be set up using any relevant definition of predicted probability of success but, as in the adaptive design modules (ADSSMod, ADTreatSel and ADPopSel functions), it will be defined using conditional power. As in these modules, conditional power is defined as the probability of establishing a significant treatment effect at the final analysis conditional upon the interim data, see Wassmer and Brannath (2016, Chapter 7). The derivation of conditional power is provided in the documentation for the ADSSMod function and will be omitted. Also, as pointed out in the adaptive design modules, the futility

stopping rules is non-binding and could be overridden by the trial's sponsor or data monitoring committee.

Let m denote the total number of doses or regimens in the trial (the doses or regimens will be referred to as treatments). Focusing on the comparison of the kth treatment versus placebo, $k = 1, \ldots, m$, conditional power for this treatment-placebo comparison is denoted by CP_k.

A treatment arm will be dropped at the interim analysis for futility if the corresponding conditional power does not exceed a pre-defined threshold denoted by c, where $0 < c < 1$. For example, the kth treatment will be dropped if

$$CP_k \leq c.$$

The trial will be terminated due to futility if all treatments are dropped.

2.3. Optimal futility stopping rule

It is common to set the futility threshold (c) to a low value such as 0.2 or 0.3. Instead of using an arbitrary threshold of this kind, an optimal value of c is easy to derive using the standard approach for defining go/no-go rules at interim looks, see, for example, Chuang-Stein et al. (2011) and Wang et al. (2014). This approach relies on the computation of the sensitivity and specificity rates for the futility stopping rule.

The sensitivity and specificity rates are defined as follows:

- The sensitivity rate is defined as the probability of correctly retaining at least one effective treatment at the interim analysis, i.e., the probability that conditional power will be greater than the futility threshold (c) in one or more treatment arms. This probability is evaluated under the alternative hypothesis of beneficial effect, i.e., all m treatments are assumed to be effective.
- The specificity rate is defined as the probability of correctly dropping all ineffective treatments at the interim analysis, i.e., the probability that conditional power will be less than the futility threshold (c) across the treatment arms. This probability is evaluated under the null hypothesis of no effect, i.e., all m treatments are assumed to be ineffective.

The sensitivity rate can be thought of as the true-positive rate and the specificity rate is equal to one minus the false-positive rate. When two or more doses or regimens (treatments) are evaluated in a trial, the definitions of the true-positive and false-positive rates account for the multiplicity of possible outcomes at this interim analysis.

Go/no-go rules are expected to demonstrate high sensitivity and specificity rates. In order to construct an optimal futility stopping rule, the futility threshold is chosen to maximize as the average of the sensitivity and specificity rates. The average of the sensitivity and specificity rates is known as the accuracy rate.

3. Case study

To illustrate the process of deriving an optimal threshold for a futility stopping rule, consider a Phase III trial in patients with schizophrenia. The efficacy profile of two doses of an experimental treatment will be evaluated compared to placebo. The primary efficacy analysis will be performed using the change from baseline to Week 6 in the PANSS (Positive and Negative Syndrome Scale) total score.

A balanced design with 160 enrolled patients per trial arm will be employed in the trial and an unblinded interim analysis will be conducted after 50% of the patients complete the 6-week treatment period or drop out of the trial prior to completing the treatment period. A 25% patient dropout rate is assumed in the trial.

To define an optimal futility stopping rule in this trial, the sensitivity rate was computed under the assumption that the true effect size in each treatment arm corresponds to the minimal clinically important difference, i.e., 0.25, and the specificity rate was computed under the assumption that the true effect size in each treatment arm is 0.

Figures 1 and 2 display the resulting sensitivity and specificity rates at the interim analysis as a function of the futility threshold (c). By definition, the sensitivity rate is a monotonically decreasing function of the threshold whereas the specificity rate is a monotonically increasing function of the threshold. An optimal value of c simultaneously increases the sensitivity and specificity rates. To find this optimal value, the accuracy rate needs to be computed.

The accuracy rate is depicted in Figure 3 and it is easy to verify that the accuracy rate is maximized if the futility threshold is set to $c = 0.23$, i.e., a treatment will be retained at the interim analysis only if conditional power is greater than 23%, otherwise it will be dropped. With this choice of the futility threshold, the sensitivity and specificity rates are both close to 80%. It also follows from Figure 3 that the curve is quite flat around the optimal value and thresholds that are reasonably close to 0.23 result in high accuracy rates. For example, the figure presents a 95% optimal interval for c, which extends from 0.04 to 0.67, and includes the thresholds for which the accuracy rate is not more than 5% worse than the optimal accuracy rate.

References

Chuang-Stein, C., Kirby, S., French, J., Kowalski, K., Marshall, S., Smith, M. K. (2011). A quantitative approach for making go/no-go decisions in drug development. Drug Information Journal. 45, 187-202.

Wang, D., Cui, L., Zhang, L., Yang, B. (2014). An ROC approach to evaluate interim go/no-go decision-making quality with application to futility stopping in the clinical trial designs. New Developments in Statistical Modeling, Inference and Application. Jin, Z., Liu, M., Luo, X. (editors). Springer, New York. 121-147.

Wassmer, G., Brannath, W. (2016). Group Sequential and Confirmatory Adaptive Designs in Clinical Trials. New York: Springer.

Figure 1. Sensitivity rate as a function of the futility threshold

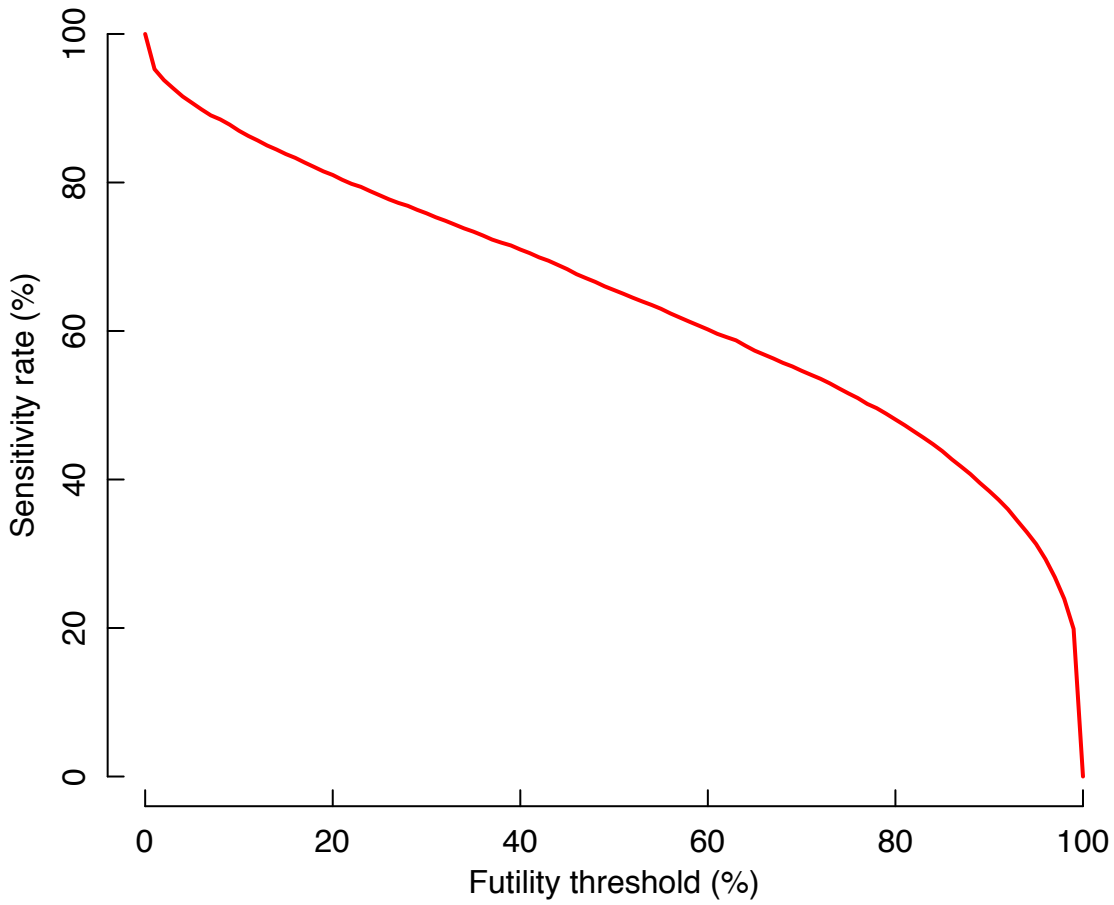

Figure 2. Specificity rate as a function of the futility threshold

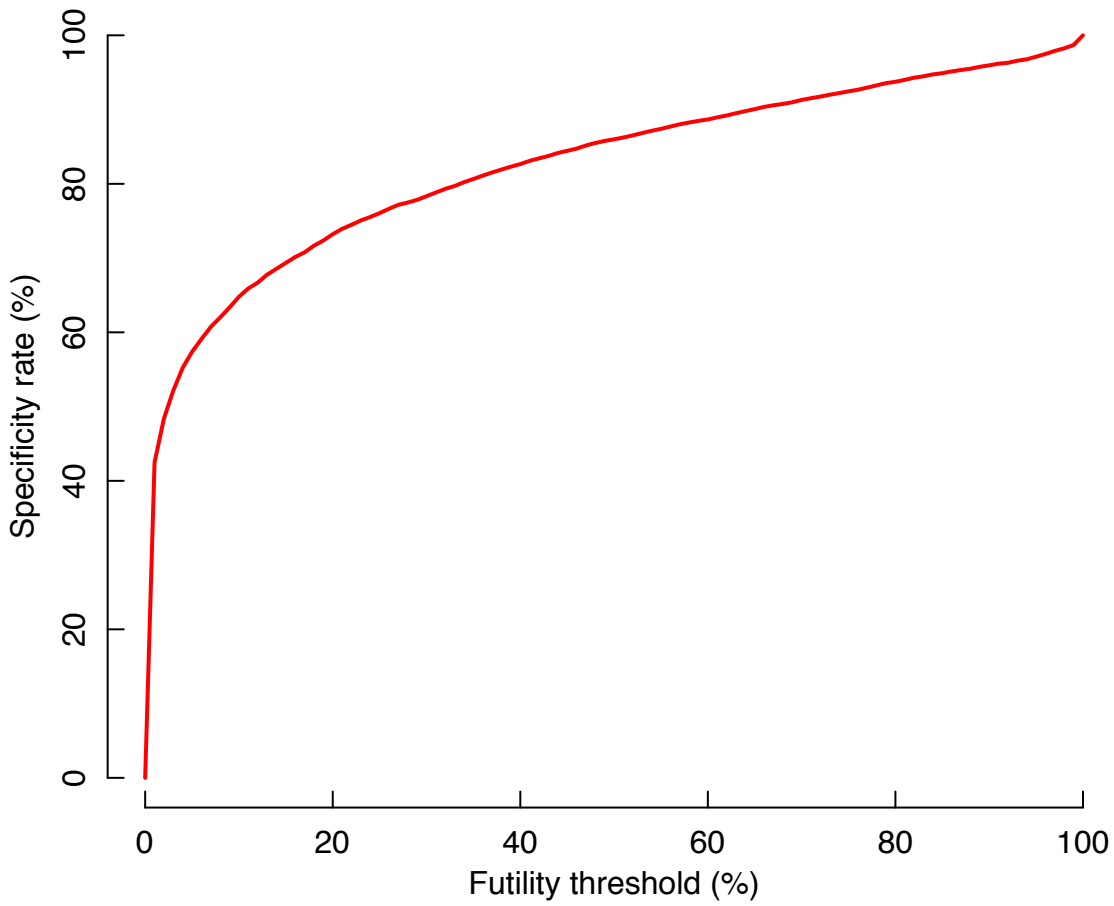

Figure 3. Accuracy rate as a function of the futility threshold