Contemporary Clinical Trials 41 (2015) 160–171

Contents lists available at ScienceDirect

Contemporary Clinical Trials

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

Selection of the treatment effect for sample size determination in a superiority clinical trial using a hybrid classical and Bayesian procedure

Maria M. Ciarleglio [a,*], Christopher D. Arendt [b], Robert W. Makuch [a], Peter N. Peduzzi [a]

[a] Department of Biostatistics, Yale School of Public Health, New Haven, CT, United States
[b] Air Force Institute of Technology, Wright Patterson AFB, OH, United States

ARTICLE INFO

Article history:
Received 22 September 2014
Received in revised form 27 December 2014
Accepted 2 January 2015
Available online 9 January 2015

Keywords:
Sample size
Clinical trial
Clinically relevant difference
Treatment effect
Hybrid classical–Bayesian

ABSTRACT

Specification of the treatment effect that a clinical trial is designed to detect (θ_A) plays a critical role in sample size and power calculations. However, no formal method exists for using prior information to guide the choice of θ_A. This paper presents a hybrid classical and Bayesian procedure for choosing an estimate of the treatment effect to be detected in a clinical trial that formally integrates prior information into this aspect of trial design. The value of θ_A is found that equates the pre-specified frequentist power and the conditional expected power of the trial. The conditional expected power averages the traditional frequentist power curve using the conditional prior distribution of the true unknown treatment effect θ as the averaging weight. The Bayesian prior distribution summarizes current knowledge of both the magnitude of the treatment effect and the strength of the prior information through the assumed spread of the distribution. By using a hybrid classical and Bayesian approach, we are able to formally integrate prior information on the uncertainty and variability of the treatment effect into the design of the study, mitigating the risk that the power calculation will be overly optimistic while maintaining a frequentist framework for the final analysis. The value of θ_A found using this method may be written as a function of the prior mean μ_0 and standard deviation τ_0, with a unique relationship for a given ratio of μ_0/τ_0. Results are presented for Normal, Uniform, and Gamma priors for θ.

1. Introduction

When designing a traditional clinical trial, the choice of the treatment effect that the trial is designed to detect (θ_A) plays a critical role in sample size determination. Under a frequentist framework, the sample size is chosen to achieve a high level of statistical power ($1-\beta$) under the alternative hypothesis ($\theta=\theta_A$) of rejecting the null hypothesis ($\theta=0$) at a given level of significance (α) [1,2]. This formal process has several subjective components, most notably being the choice of θ_A. In part, this subjectivity is due to varying definitions for θ_A [3,4]. Some take θ_A to represent the minimum clinically significant/important difference, or the smallest treatment effect that one would accept to claim superiority of the new treatment relative to the control. Others take θ_A to represent an expected difference, often based on a treatment effect that has been previously demonstrated in similar agents or in early phase studies of the current therapy [5,6]. This dichotomy has led to a natural merging of the two definitions and, when such an intersection exists, θ_A equals both a realistic and important effect [7]. Unfortunately, the choice of θ_A is often easily swayed, with the difficulty of finding and enrolling the required sample size having more influence than considerations for what is

* Corresponding author at: Department of Biostatistics, Yale School of Public Health, 60 College Street, New Haven, CT 06510, United States. Tel.: +1 203 737 4829.
E-mail addresses: maria.ciarleglio@yale.edu (M.M. Ciarleglio), christopher.arendt@us.af.mil (C.D. Arendt), robert.makuch@yale.edu (R.W. Makuch), peter.peduzzi@yale.edu (P.N. Peduzzi).

http://dx.doi.org/10.1016/j.cct.2015.01.002
1551-7144/

realistic and important. The purpose of this paper is to provide quantitative guidance when choosing a value of θ_A based on prior evidence about the true measure of association, θ.

In practice, the choice of the treatment effect is often based on the effect seen in prior studies. Meta-analysis is frequently used to combine the effects from each study into an average effect. These results, together with clinical opinion, establish the hypothesized value of the treatment effect in the current trial, and traditional sample size formulas incorporate a belief about θ through this single point estimate. However, once the point estimate is determined, the variability between the effects measured in previous studies and the uncertainty in clinical beliefs about the effect are neglected in the traditional formulation of power. Simply projecting the power of a planned study based on the observed maximum likelihood estimator from a previous experiment [8] or other empirical estimator [9] tends to overestimate the power of the planned study and does not properly account for the uncertainty in the hypothesized treatment effect.

We present a "hybrid classical and Bayesian" [7] technique based on conditional expected power [10] that incorporates the distribution of available prior information on θ from historical data and clinical opinion to determine a reasonable estimate of the treatment effect that can be detected with a desired power in a superiority clinical trial. Unlike traditional power, which is calculated assuming the truth of a point alternative hypothesis ($\theta = \theta_A$), conditional expected power conditions on the truth of a composite alternative of superiority (e.g., $\theta > 0$). Using conditional expected power allows us to formally integrate prior information into the determination of sample size by averaging traditional frequentist power curves using the conditional prior distribution of θ as the averaging weight. The value of θ that yields conditional expected power equal to the desired level of frequentist power is then used as the value of θ_A in traditional sample size calculations. If the value of θ_A falls below what is considered to be clinically relevant, the results of this method will encourage serious discussion concerning the feasibility of the study.

Prior work done in this area [3,9–16] recognizes the importance of accounting for the uncertainty in the study parameters when designing a clinical trial. The literature describes how different measures of prior-adjusted power (discussed in Section 2) can provide insight into the design and proposes directly finding the sample size n that gives the desired level of power. We generalize the Lan and Wittes [16] implementation of Brown's [10] method for a truncated prior. This method maintains the traditional notion of power as being conditional on the truth of the alternative hypothesis of superiority. Rather than directly finding n, however, we frame the technique to provide a solution for θ_A to maintain the traditional setting for determining sample size. By developing an explicit procedure for using information on the past performance and expert opinion regarding future performance of a treatment when considering the power of the current study, we hope to encourage a more rigorous and substantiated choice of θ_A that will lead to a rational discussion of the feasibility of the study and reduce the use of overly large treatment effects that are not supported by historical data. We illustrate this technique using three prior distributions for θ: a Normal prior, a Uniform prior, and a Gamma prior and show that there is a unique relationship between the mean and standard deviation of the prior distribution and the value of θ_A that gives the desired level of conditional expected power.

2. Power

2.1. Traditional power

In many situations, it is reasonable to assume a Normal or approximate Normal sampling distribution for the summary statistic calculated based on n observations, $Y_n \sim N(\theta, \sigma/\sqrt{n})$. The parameter of interest θ represents a treatment effect such as the difference in the mean response between two groups for the case of continuous data, the risk difference or the log odds ratio for binary data, the log hazard ratio for survival data, or the log rate ratio for count data [4].

Assume that after n observations, a frequentist significance test is performed to test the null hypothesis $H_0 : \theta = 0$ versus the alternative hypothesis $H_1 : \theta \neq 0$, where $\theta > 0$ indicates benefit of the experimental treatment. In this setting, H_0 is rejected at the α-level if $|y_n| > -z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, where y_n is the sample statistic and $z_{\alpha/2}$ is the standard normal deviate for tail area $\alpha/2$. A positive conclusion, D_1, occurs if $y_n > -z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. The traditional power for this test to detect a difference of magnitude $\theta = \theta_A$ is $P(D_1|\theta = \theta_A) = \Phi\left[z_{\alpha/2} + \theta_A\sqrt{n}/\sigma\right]$, where $\Phi[\cdot]$ is the standard normal cumulative distribution function. Traditional power is often set to a target level of $1 - \beta$. Since we are interested in detecting a successful outcome, we only consider the traditional power curve that gives probabilities associated with a positive conclusion. In a two-sided test, the contribution to statistical power in the direction of treatment harm, $\Phi\left[z_{\alpha/2} - \theta_A\sqrt{n}/\sigma\right]$, is negligible.

Sample size is chosen that gives a high level of traditional power (e.g., $1 - \beta = 0.8$ to 0.9) to detect an effect at least of size θ_A for known variance, σ^2, yielding the traditional sample size formula,

$$n = \frac{\left(z_{1-\beta} - z_{\alpha/2}\right)^2\sigma^2}{\theta_A^2}. \tag{1}$$

The probability of detecting a statistically significant difference conditional on a hypothesized value for θ under H_1 is typically plotted as a function of θ,

$$P(D_1|\theta) = \Phi\left[z_{\alpha/2} + \frac{\theta\sqrt{n}}{\sigma}\right], \tag{2}$$

giving the traditional power curve.

2.2. Expected power and conditional expected power

The traditional power curve does not account for the uncertainty associated with the treatment effect θ and does not indicate if the planned sample size is adequate given this uncertainty. The idea of "average power" was introduced as a way to use the distribution of prior beliefs about the treatment effect to provide an overall predictive probability of a positive conclusion [3]. The *expected power* (EP), also known as average power [4] or assurance [14], averages the traditional power curve, $P(D_1|\theta)$, over the plausible range of θ using the prior