Contemporary Clinical Trials 32 (2011) 561–568

Contents lists available at ScienceDirect

Contemporary Clinical Trials

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

A comparison of two worlds: How does Bayes hold up to the status quo for the analysis of clinical trials?

Alice R. Pressman *, Andrew L. Avins, Alan Hubbard, William A. Satariano

ARTICLE INFO

Article history:
Received 20 December 2010
Accepted 15 March 2011
Available online 29 March 2011

Keywords:
Clinical trials
Bayesian analysis
Frequentist analysis
Group sequential analysis
Analysis of clinical trials
Clinical trials methodology

ABSTRACT

Background: There is a paucity of literature comparing Bayesian analytic techniques with traditional approaches for analyzing clinical trials using real trial data.
Methods: We compared Bayesian and frequentist group sequential methods using data from two published clinical trials. We chose two widely accepted frequentist rules, O'Brien–Fleming and Lan–DeMets, and conjugate Bayesian priors. Using the nonparametric bootstrap, we estimated a sampling distribution of stopping times for each method. Because current practice dictates the preservation of an experiment-wise false positive rate (Type I error), we approximated these error rates for our Bayesian and frequentist analyses with the posterior probability of detecting an effect in a simulated null sample. Thus for the data-generated distribution represented by these trials, we were able to compare the relative performance of these techniques.
Results: No final outcomes differed from those of the original trials. However, the timing of trial termination differed substantially by method and varied by trial. For one trial, group sequential designs of either type dictated early stopping of the study. In the other, stopping times were dependent upon the choice of spending function and prior distribution.
Conclusions: Results indicate that trialists ought to consider Bayesian methods in addition to traditional approaches for analysis of clinical trials. Though findings from this small sample did not demonstrate either method to consistently outperform the other, they did suggest the need to replicate these comparisons using data from varied clinical trials in order to determine the conditions under which the different methods would be most efficient.

1. Introduction

The high price tag associated with clinical trials has motivated researchers to find more cost-effective ways to conduct research. In addition, research ethics mandate that we do not continue clinical studies beyond the point at which sufficient information is available to answer the research question, so that the smallest number of patients receive the inferior therapy [1,2]. Finally, it is important to complete clinical trials as quickly as possible to assure that superior treatments are incorporated into regular practice in a timely fashion.

Bayesian analysis is a widely promoted response to these mandates. With Bayesian methods, prior knowledge is formally organized to direct the course of the study, and participants are randomized only as necessary [3,4]. Though these methods are increasingly accepted in the field of cancer treatment and device trials [5–11], they are seldom implemented in other types of trials.

Traditional ("frequentist") statistical methods have evolved to include group sequential analysis (in which a series of interim analyses is performed over the collection of the sample) [12]. This technique allows for the early termination of a trial, and thus can save funds, participant time, and provide knowledge about therapeutic efficacy more quickly. Frequentist conclusions rely on the preservation of an experiment-wise error rate (alpha level), and the size of this error relies heavily

* Corresponding author at: Kaiser Permanente Division of Research, 2000 Broadway, Oakland, CA 94612, United States. Tel.: +1 510 891 3236; fax: +1 5108913606.
E-mail address: alice.pressman@kp.org (A.R. Pressman).

1551-7144/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.cct.2011.03.010

on the fact that the data be processed only once. Because this issue of multiple testing is tantamount, frequentist statisticians have developed alpha spending functions which penalize the alpha level at each look based on the number of interim analyses and the amount of accumulated data [13–18].

While frequentist methodology is the widely accepted paradigm, Bayesian methodology has not, until recently, been easily accessible because of its need for enormous computer resources. This reliance on high-level computer programming combined with the perceived subjective nature of the priors has resulted in skepticism from clinical investigators.

Statisticians have demonstrated the efficiency of Bayesian methods with simulated data under various scenarios [4,6,19,20], but many clinical researchers remain unconvinced for several reasons. First, simulations can be constructed to favor one's preferred methods, and though many articles in the literature demonstrate Bayesian advantages in real observational data, we have found only three such demonstrations of the comparison between Bayesian and frequentist methods for the analysis of clinical trials using real data [21–23]. Thus, these clinical-trial methods need to be investigated using real data where one can examine their performance with regards to criteria researchers and policy makers care about. Second, simulations are inevitably based on assumptions which may not be realistic. Finally, the status quo demands meeting certain operating criteria, such as Type I error rates and power, which are frequentist concepts not inherent in Bayesian methods.

We compared these two approaches with data from completed clinical trials in order to test their performance in the real world. Our goal was to use a simple methodology to compare the results under Bayesian and frequentist analyses on the same data. To that end, we performed *post hoc* analyses of the data from two completed clinical trials.

2. Analytical methods

2.1. Bayesian

In the context of clinical trials, Bayesian analysis is an iterative process in which investigators use all available data (external evidence) and prior knowledge to construct a prior distribution of the parameter of interest (e.g., the between-group difference in treatment outcomes). Next, part of an experiment is conducted and the results (called the likelihood) are applied to the prior distribution to obtain an updated "posterior" distribution. This posterior distribution is then used to calculate the probability that the treatment is superior to the control. If these posterior probabilities dictate the continuation of the data collection process, the posterior distribution serves as the prior distribution for the next iteration [4] (see Appendix).

2.2. Frequentist

Frequentist statistical methodology depends on the assumption that sampled data come from a population with a specific distribution and set parameters. Further, the frequentist concept of probability is based on long-run expected frequencies of occurrence. Data are collected and used to test hypotheses about the value of the fixed parameter. The goal is to use estimates of variation in repeated experiments to determine whether the observed data are consistent with a specified null distribution [24].

Clinicians apply Bayesian principles to the science of diagnostic testing by incorporating likelihood and prior probability into recommendations they present to patients [25]. However, few clinical trials use Bayesian methods, primarily because the techniques are difficult to understand, and depend largely on the investigator's specific assumptions. In addition, the performance of these techniques, evaluated within the frequentist world of gold standards (e.g., type I and II error rates) has not been adequately addressed [8,11,26].

Bayesian and frequentist methods differ in their goals. Bayesian methods seek to determine the probability that the population has a certain characteristic, given the observed data and the prior information, whereas frequentist methods seek to determine the probability that we would see the observed data if the null hypothesis were true. Put another way, both deal with conditional probabilities, but frequentist inference centers around P(data|parameter) while Bayesian is concerned with P(parameter|data) [23]. The frequentist approach to an effectiveness trial is to choose a natural null (usually no effect), and examine whether the data can provide evidence against it, whereas the Bayesian approach is to choose a hypothesis about the presence of an effect and assess evidence in its favor.

Our primary goal was to compare frequentist group sequential and Bayesian clinical trial analyses to determine how sensitive Bayesian methodology is to starting assumptions, as measured by trial outcome under different prior distributions, as practically applied to actual trial data. We considered both informative (i.e. distributions in which we used prior knowledge to inform the initial parameter estimates) and non-informative priors (i.e. flat prior distributions that assume no prior knowledge regarding the superiority of either treatment arm). A secondary aim was to examine how frequentist group sequential analysis differs by method of conservation of Type I error as measured by the outcome in each trial under two commonly used frequentist group sequential methods: Obrien–Fleming, and Lan–DeMets power function. Finally, we combined the results to determine what proportion of Bayesian sequential analysis methods would yield different outcomes from their frequentist counterparts in a series of comparative simulations.

3. Study methods

3.1. The data

We considered a variety of trial characteristics including sample size, number of events, type of outcome, availability of datasets and relative impact in their fields. We present the re-analysis of two such trials to represent two distinct types of outcomes (continuous and time-to-event).

The Studies of Left Ventricular Dysfunction–Treatment Trial (SOLVD-TT) was a double-blind, placebo-controlled trial of enalapril in patients with symptomatic heart failure with a primary endpoint of all-cause mortality [27]. The study followed 2569 participants for a mean of 41 months; final results showed a 16% reduction in total mortality among the enalapril-allocated participants (HR = 0.84, 95% CI: 0.74, 0.95) [27,28]. Data for these analyses were obtained from the