Contemporary Clinical Trials 45 (2015) 103–112

Contents lists available at ScienceDirect

Contemporary Clinical Trials

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

Review

Design of clinical trials with failure-time endpoints and interim analyses: An update after fifteen years

Pei He [a,*], Tze Leung Lai [b], Zheng Su [c]

[a] Genentech Inc., South San Francisco, CA 94080, USA
[b] Department of Statistics, Stanford University, Stanford, CA 94305, USA
[c] Deerfield Institute, New York, NY 10017, USA

ARTICLE INFO

Article history:
Received 7 May 2015
Received in revised form 25 May 2015
Accepted 26 May 2015
Available online 29 May 2015

Keywords:
Adaptive design
Calendar time
Early stopping
Multiple endpoints
Nonproportional hazards
Survival analysis

ABSTRACT

Time to event is the clinically definitive endpoint in Phase III trials of new treatments of cancer, cardiovascular and many other diseases. Because these trials involve relatively long follow-up, their protocols usually incorporate periodic interim analyses of the data by a Data and Safety Monitoring Board/Committee. This paper gives a review of the major developments in the design of these trials in the 21st century, spurred by the need for better clinical trial designs to cope with the remarkable advances in cancer biology, genomics and imaging that can help predict patients' sensitivity or resistance to certain treatments. In addition to this overview and discussion of related issues and challenges, we also introduce a new approach to address some of these issues.

Contents

1. Introduction

Analysis of clinical studies with failure-time endpoints has been an important topic in biostatistics and has also led to a number of major methodological advances and important breakthroughs in statistical

* Corresponding author.
E-mail address: he.pei.hepei@gmail.com (P. He).

http://dx.doi.org/10.1016/j.cct.2015.05.018
1551-7144/

theory. A celebrated example is Cox's proportional hazards regression [1] that led to subsequent developments in partial likelihood, semiparametric efficiency, and statistical analysis of counting processes [2–6]. Although less renowned in comparison, the design of clinical trials with failure-time endpoints has also had important impact on clinical trial biostatistics and led to innovations in data monitoring and interim analysis of clinical trials. These innovations dated back to the seminal papers [7,8] in 1982 on the Beta Blocker Heart Attack Trial (BHAT) and have continued until today, although at a much less spectacular pace than survival analysis. In this paper we give a review of the major developments in the 21st century, hence the "fifteen years" in the title. The "update" in the title refers to updating a previous review [9] that also provides a computer program to determine the power and sample size in the trial design; note that "design of clinical trials with failure-time endpoints and interim analyses" in our title is also the main part of the title of [9]. Since last year, we and our colleagues at Stanford University's Center for Innovative Study Design have been working to develop open-source software, which can be considered as an update of [9] for the implementation of some of the methods described in the next two sections. Section 2 reviews several new trends and innovative methods after the publication of [9] in the predecessor of this journal. In particular, it describes hybrid resampling for valid inference on primary and secondary endpoints of a survival trial, choice of stopping rules, and adaptive designs including seamless Phase II–III designs.

In his 2010 budget request, the Director of the National Cancer Institute earmarked "re-engineering" cancer clinical trials as a research initiative. The reason why re-engineering is needed is that although remarkable progress in biomedical sciences raised new hope for cancer treatment, the hope did not materialize because of the relatively small number of new anticancer agents that were demonstrated to be efficacious in Phase III clinical trials, for which time to event (typically overall survival and occasionally progression-free survival) is a definitive endpoint. Besides choice of stopping boundaries [9], also considers choice of test statistics. Being able to choose appropriate test statistics at terminal analysis can substantially increase the power of the commonly used logrank statistics in current designs that are mostly based on hazard ratios of treatment to control. In Section 3 we develop a new approach that allows adaptive choice of the test statistics at terminal analysis while still maintaining the prescribed Type I error. This circumvents one of the widely recognized difficulties with current survival trial designs that are dominated by hazard ratios and logrank statistics, which are inefficient for nonproportional hazards. The year 2010 also marked the appearance of the much awaited FDA Draft Guidance for Industry on Adaptive Design. Two years later, the President's Council of Advisors on Science and Technology (PCAST) issued a report on "Propelling Innovations in Drug Discovery, Development, and Evaluation" and argued for using "innovative new approaches for trial design that can provide more information more quickly" as "it is increasingly possible to obtain clear answers with many fewer patients and with less time" by focusing studies on "specific subsets of patients most likely to benefit, identified based on validated biomarkers." Section 4 begins with a review of ongoing work in this direction for drug development and confirmatory testing, some of which is related to the approach introduced in Section 3. It then proceeds with further discussion and several concluding remarks.

2. Stopping rules, adaptive designs, and hybrid resampling

In Section 2.1 we review developments in the choice of stopping rules for time-sequential survival trials, in which "survival" refers to the failure-time endpoint in the title and "time-sequential" encapsulates the "interim analyses" that are carried out at prespecified calendar times. As pointed out in [10], survival trials have two timescales — calendar time t and information time $V(t)$, which is the null variance of the test statistic at t. The information time $V(t)$ is the intrinsic time-scale for interim data but is typically unknown before time t unless restrictive assumptions are made *a priori*. In the past fifteen years, seamless Phase II–III designs and Bayesian adaptive designs are the most active area of research in innovative clinical trial designs. Section 2.3 gives a review of some of these developments in the context of time-sequential survival trials. Another important development in this period is hybrid resampling [11–13], which is reviewed in Section 2.2 and provides a basic tool in the methodological development in Section 3.

2.1. Early stopping for efficacy or futility at interim analysis

This basic problem in time-sequential survival trials is already addressed in [9], and we describe here subsequent developments. The censored rank statistics considered in [9] and its precursors [14,15] have the general form

$$S_n(t)=\sum_{i=1}^{n}\delta_i(t)\psi\left(H_{n,t}(X_i(t))\right)\left\{1-\frac{m_{n,t}(X_i(t))}{m'_{n,t}(X_i(t))+m''_{n,t}(X_i(t))}\right\} -\sum_{j=1}^{n'}\delta'_j(t)\psi\left(H_{n,t}(Y_j(t))\right)\frac{m'_{n,t}(Y_j(t))}{m_{n,t}(Y_j(t))+m'_{n,t}(Y_j(t))}, \tag{1}$$

where ψ is a nonrandom continuous function on $[0,1]$, $n=n'+n''$ is the total sample size, with n' patients assigned to treatment X and n'' assigned to treatment Y, $m'_{n,t}(s)=\sum_{i=1}^{n'}I(X_i(t)\geq s)$, $m''_{n,t}(s)=\sum_{j=1}^{n''}I(Y_j(t)\geq s)$, and $X_i(t)$, $Y_j(t)$, $\delta_i'(t)$, $\delta_j''(t)$ and $H_{n,t}(\cdot)$ are defined below. Let $T_i'\geq 0$ denote the entry time and $X_i>0$ the survival time (or time to failure) after entry of the ith subject in treatment group X and let T_j'' and Y_j denote the entry time and survival time after entry of the jth subject in treatment group Y. The subjects are followed until they fail or withdraw from the study or until the study is terminated. Let ξ_i' (ξ_j'') denote the time to withdrawal, possibly infinite, of the ith (jth) subject in the treatment group X (Y). Thus the data at calendar time t consist of $(X_i(t),\delta_i'(t))$, $i=1,\ldots,n'$, and $(Y_j(t),\delta_j''(t))$, $j=1,\ldots,n''$, where $X_i(t)=\min(X_i,\xi_i',(t-T_i')^+)$, $\delta_i'(t)=I(X_i(t)=X_i)$, and $Y_j(t)$ and $\delta_j''(t)$ are defined similarly in terms of Y_j, ξ_j'' and T_j''. Let $H_{n,t}$ be the left-continuous version of the Kaplan–Meier estimator of the distribution function of the combined sample, defined by

$$1-H_{n,t}(s)=\prod_{u<s}\left\{1-\frac{\Delta N'_{n,t}(u)+\Delta N''_{n,t}(u)}{m'_{n,t}(u)+m''_{n,t}(u)}\right\}, \tag{2}$$

where $N'_{n,t}(s)=\sum_{i=1}^{n'}I\left(X_i\leq\xi_i'\wedge(t-T_i')^+\wedge s\right)$, $N''_{n,t}(s)=\sum_{j=1}^{n''}I(Y_j\leq\xi_j''\wedge(t-T_j'')^+\wedge s)$, $\Delta N(s)=N(s)-N(s-)$ and we use the convention $0/0=0$. For the time-sequential censored rank statistics Eq. (1), Gu and Lai [14] showed that $\{S_n(t)/\sqrt{n},t\geq 0\}$ converges weakly to a Gaussian process with independent increments and variance function $V(t)$ under the null hypothesis $H_0:F=G$ and contiguous alternatives. Two commonly used estimates of $V(t)$ are

$$V_n^1(t)=\int_0^t\frac{\psi^2(H_{n,t})(s)m'_{n,t}(s)m''_{n,t}(s)}{\left(m'_{n,t}(s)+m''_{n,t}(s)\right)^2}\,d\left(N'_{n,t}(s)+N''_{n,t}(s)\right), \tag{3}$$

or

$$V_n^2(t)=\int_0^t\frac{\psi^2(H_{n,t}(s))}{\left(m'_{n,t}(s)+m''_{n,t}(s)\right)^2}\left\{\left(m''_{n,t}(s)\right)^2dN'_{n,t}(s)+\left(m'_{n,t}(s)\right)^2dN''_{n,t}(s)\right\}. \tag{4}$$

As a compromise between these two choices, Gu and Lai [14] also considered $V_n^3(t)=\{V_n^1(t)+V_n^2(t)\}/2$. For any choice $V_n(t)$ of the three estimates, $n^{-1}V_n(t)$ converges in probability to $V(t)$ under H_0 and