Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computational Statistics & Data Analysis 49 (2005) 217–226
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Equivalence of two treatments and sample size determination under exponential survival model with censoring

Jun-mo Nam[a], Jinheum Kim[b], Seungyeoun Lee[c,*]

[a]*Department of Health and Human Services, Biostatistics Branch, DCEG, National Cancer Institute, National Institutes of Health, Rockville, MD 20852-7240, USA*
[b]*Department of Applied Statistics, University of Suwon, Gyeonggi-Do 445-743, Republic of Korea*
[c]*Department of Applied Mathematics, Sejong University, 98 Gunja-dong, Gwangjin-gu, Seoul, 143-747, Republic of Korea*

Received 9 February 2004; received in revised form 30 April 2004

Abstract

We present the likelihood score and F statistics for ascertaining equivalence of two treatments in survival under an exponential model with independent censoring. We provide explicit formulae for power and sample size requirement for trials using the score and F tests, and compare the score and F tests with the log rank test by Com-Nougue et al. (Statist. Med. 12 (1993) 1353). Simulation results show that empirical powers of the score, F and log rank tests are satisfactorily close to the corresponding asymptotic powers for small-to-moderate sample size. We find these three methods are essentially identical in terms of level and power. However, the score and F methods are very sensitive to departure from the exponential assumption while the log rank test is more robust. The methods are illustrated by application to data from a randomized trial of two treatments for B non-Hodgkin lymphoma. Published by Elsevier B.V.

Keywords: Censoring; Equivalence trial; Exponential model; Score test; F test; Power and sample size

1. Introduction

Recently, there has been a great interest in establishing equivalence of two treatments in clinical trials. For example, a standard chemotherapy in pediatric oncology is highly

* Corresponding author. Tel.: +82-2-3408-3161; fax: +82-2-3408-3315.
E-mail addresses: namj@mail.nih.gov (J.-m. Nam), jinhkim@suwon.ac.kr (J. Kim), leesy@sejong.ac.kr (S. Lee).

0167-9473/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.csda.2004.05.009

effective but causes severe toxic side-effects, and researchers are interested in a less toxic new treatment which may be essentially as effective as the standard one in survival (Patte et al., 1991). The intention of an equivalence trial is to demonstrate that two treatments do not differ by more than a prescribed small amount which is materially insignificant. The conventional test procedure for detecting a difference in a comparative trial cannot be applied for this situation.

Statistical methods for establishing one-sided equivalence or non-inferiority of a new treatment to the standard one on binary responses have been investigated by many authors, e.g., Dunnett and Gent (1977), Roebruck and Kühn (1995) and Nam (1997). For equivalence of two survival distributions with censored observations, Wellek (1993) and Com-Nougue et al. (1993) have proposed testing procedures based on the proportional hazards model. Wellek (1993) derived the uniformly most powerful test in terms of the maximum partial likelihood estimator but the sample size equation is not given explicitly while Com-Nougue et al. (1993) provided the confidence intervals for the actual hazard ratio based on the log rank test statistic. When data follow an exponential model with no censoring, Bristol and Desu (1990) have suggested a parametric method of testing for equivalence. However, a parametric method based on censored data has not been thoroughly studied.

In this paper, we investigate statistical methods involving the equivalence of two treatments based on exponentially distributed survival data with censoring. In Section 2, we derive two different tests for equivalence: the score test and F test procedures. In addition, the asymptotic powers and approximate sample size formula are provided. In Section 3, the score and F tests are compared with the log rank test (Com-Nougue et al., 1993) by simulations in level and power, and approximate numbers of events required for a specific power using these methods are examined. Also, we investigate the robustness of the three tests when the underlying exponential model is violated. Sections 4 and 5 contain an example based on non-Hodgkin's malignant type B lymphoma data and discussion.

2. Test statistics and power functions

2.1. Score method

Denote the survival and censoring times of standard and new treatment groups by t_{ij} and c_{ij} for $i = 0, 1$ and $j = 1, 2, \ldots, n_i$, respectively. The first subscript $i = 0, 1$ indicates the standard and new treatment groups and the second subscript indicates the jth individual in the ith group. Under right censorship, we observe survival data $\{(x_{ij}, \delta_{ij}), i = 0, 1$ and $j = 1, 2, \ldots, n_i\}$, where $x_{ij} = \min(t_{ij}, c_{ij})$ and $\delta_{ij} = I(t_{ij} \leqslant c_{ij})$, i.e., $\delta_{ij} = 1$ if $t_{ij} \leqslant c_{ij}$ and $\delta_{ij} = 0$ otherwise. Assume that t_{ij} and c_{ij} are independent within a group.

Consider the exponential survival distributions of the standard and the new treatment groups as $S_0(t) = \exp(-h_0 t)$ and $S_1(t) = \exp(-h_1 t)$, where $h_i > 0$ for $i = 0, 1$. Denote the hazard ratio by $r = h_1/h_0$. Let $x_{i\cdot} = \sum_{j=1}^{n_i} x_{ij}$ and d_i be total survival follow-up time and the number of uncensored observations for $i = 0, 1$, respectively. The score statistic for testing $H_0 : r \geqslant r_0$ against $H_1 : r < r_0$ can be simplified as

$$z = \{d_1 - \widehat{\mu}(r_0)\}/\{\widehat{v}(r_0)\}^{1/2}, \tag{1}$$