Contemporary Clinical Trials 33 (2012) 104–115

Contents lists available at SciVerse ScienceDirect

Contemporary Clinical Trials

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

A comparison of statistical approaches for physician-randomized trials with survival outcomes

Margaret R. Stedman [a,b,*], Robert A. Lew [c,d], Elena Losina [a,b,c], David R. Gagnon [c,d], Daniel H. Solomon [e,b], M. Alan Brookhart [f]

[a] Orthopedics and Arthritis Center for Outcomes Research, Department of Orthopedics, Brigham and Women's Hospital, Boston, MA, USA
[b] Harvard Medical School, Boston, MA, USA
[c] Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA
[d] MAVERIC, VA Cooperative Studies, Boston V.A. Healthcare System, Boston, MA, USA
[e] Division of Rheumatology, Immunology, Allergy, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA
[f] Department of Epidemiology, University of North Carolina Gillings School of Global Public Health, Chapel Hill, NC, USA

ARTICLE INFO

Article history:
Received 1 April 2011
Revised 17 August 2011
Accepted 31 August 2011
Available online 6 September 2011

Keywords:
Cluster randomized trials
Survival analysis
Physician-randomized trials
Permutation test
Simulation study
Shared frailty model

ABSTRACT

This study compares methods for analyzing correlated survival data from physician-randomized trials of health care quality improvement interventions. Several proposed methods adjust for correlated survival data; however the most suitable method is unknown. Applying the characteristics of our study example, we performed three simulation studies to compare conditional, marginal, and non-parametric methods for analyzing clustered survival data. We simulated 1000 datasets using a shared frailty model with (1) fixed cluster size, (2) variable cluster size, and (3) non-lognormal random effects. Methods of analyses included: the nonlinear mixed model (conditional), the marginal proportional hazards model with robust standard errors, the clustered logrank test, and the clustered permutation test (non-parametric). For each method considered we estimated Type I error, power, mean squared error, and the coverage probability of the treatment effect estimator. We observed underestimated Type I error for the clustered logrank test. The marginal proportional hazards method performed well even when model assumptions were violated. Nonlinear mixed models were only advantageous when the distribution was correctly specified.

Published by Elsevier Inc.

1. Introduction

Physician-randomized trials are clinical trials where each physician is randomized to a treatment or control intervention. Although the intervention is delivered at the physician level, effectiveness of the intervention is often measured at the patient level, so that patient data are clustered by the randomized physician. This serves to reduce contamination among participants and allows physicians to provide consistent treatment to all their patients. However, the design induces correlations in the data which must be accounted for in the analysis. Performance of these methods depends upon: the number of clusters, the cluster size, and the correlation structure. Typically the correlation structure is unknown, the number of clusters exceeds one hundred, and the cluster sizes vary greatly [1].

Usual methods of analysis for survival data: the logrank test, the Cox proportional hazards model, and the accelerated failure time model, assume the data are independent. If these methods are applied to correlated data the estimation of effect size may be correct, but standard errors will not be adjusted for the correlations in the data and Type I error will exceed nominal size. Several methods have been proposed to correct the standard errors in these analyses. Jung and Jeong recently published a method for the clustered logrank test [2]. Lee, Wei, and Amato [3] popularized the marginal

* Corresponding author at: Orthopedics and Arthritis Center for Outcomes Research, Department of Orthopedics, Brigham and Women's Hospital, 75 Francis Street, BC-4-4016, Boston, MA 02115, USA. Tel.: +1 617 525 7973.
E-mail addresses: mstedman2@partners.org, mstedman2@gmail.com (M.R. Stedman).

1551-7144/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.cct.2011.08.008

proportional hazards model with robust standard errors. Alternatively, one could model the frailty with a nonlinear mixed model as a counterpart to the parametric accelerated failure time approach [4].

Few studies have attempted to compare methods for analyzing clustered survival data. Glidden et al. [5] performed a simulation study of a clustered design where randomization occurred within cluster and found that conditional models performed better than marginal models. Loeys et al. [6] compared marginal and conditional methods for cluster randomized trials and found that the power was similar between the marginal and conditional methods. Cai et al. [7] compared the clustered permutation test to the usual logrank test and found that the clustered permutation test improved preservation of Type I error.

In this simulation study we compare marginal, conditional, and nonparametric methods for clustered survival analysis under the conditions of fixed cluster size, variable cluster size, and various random effects. We examine power, Type I error, bias, coverage, and mean squared error to determine the best method of analysis for physician-randomized trials. All data are simulated using the design and characteristics of two physician-randomized trials of an educational intervention for osteoporosis management. In Section 2 we begin with a description of two example studies. In Section 3 we discuss the application of the shared frailty model to the physician-randomized trial. In Section 4 we review the various estimation methods that may be used to analyze the data. In Sections 5 and 6 we perform a simulation study to evaluate the methods. Actual results from the example studies are presented for comparison in Section 6.2. In Section 7 we make recommendations on how best to analyze physician-randomized trials based on this research.

2. Example studies

Osteoporosis is a disease of the elderly that makes bones prone to fracture. According to practice guidelines, all patients at moderate to high risk of osteoporosis should either receive a bone mineral density scan to rule out osteoporosis or be prescribed a preventive medication to treat the disease [8]. Despite these guidelines many patients do not receive adequate treatment [8]. Physician education or "academic detailing" programs have been designed to improve management and prevention of this disease, however the effectiveness of this education effort as an intervention is unknown. We attempted to evaluate the effectiveness of the education program on improving osteoporosis management in two trials, one occurring in Pennsylvania, the other in New Jersey [9,10]. Data were analyzed by survival analysis to determine if the physician education program significantly improved a patient's chances of being correctly managed to prevent disease. Both trials served as models in the design of the simulation studies.

The PACE study enrolled Pennsylvania Medicare beneficiaries who were eligible for a state-run pharmaceutical benefits plan. Patients at risk of osteoporosis were selected for the physician-randomized trial with a two-way factorial design. A total of 828 physicians with 13,455 patients participated in one of four groups: 1) physician education, 2) patient education, 3) patient education plus physician education, or 4) usual care [9]. For the purposes of this simulation we focused on comparing two groups: group 3, patient education plus physician education and group 4, the usual care group. The average age of this study population was 82. Patient claims data were used to measure the outcome of a bone mineral density scan or prescription for a preventative osteoporosis medication. Patients were followed using insurance claims data for 487 days or until they lost insurance coverage, were admitted to a nursing home, or died. From this, approximately 10% of the patients were censored. On average there were 16 patients per physician with a range of 2 to 65 patients per physician. See Fig. 1 for distribution of the number of patients per physician.

The HORIZON study enrolled patients at-risk of osteoporosis from the Horizon New Jersey health insurance plan. This population was slightly younger than the PACE study with an average age of 68. Loss due to lack of coverage, nursing home admittance, or death was 15%. The trial randomized 434 physicians with a total of 1973 patients to a combination treatment of physician and patient education. Patients were followed using patient claims data to determine if they received osteoporosis management (bone mineral density scan or preventive osteoporosis medication). On average there were four patients per physician with a range of 1 to 148 patients per physician (see Fig. 1) [10].

3. The model

We used a shared frailty model to represent the example physician-randomized trials [11]. We defined $k=1,\ldots,K$ physician clusters assigned at random to either treatment ($X=1$) or control ($X=0$) group. Each physician contributed $i=1,\ldots,n_k$ patients to the study so that $N=\sum_{k=1}^{K} n_k$ was the total number of patients in the study. Let $(T_{ki}; k=1,\ldots,K, i=1,\ldots,n_k)$ be the time until each patient received either a bone mineral density scan or preventive osteoporosis medication. Let $(D_{ki}; k=1,\ldots,K, i=1,\ldots,n_k)$ be the censoring time for patient i of physician k. $X_k; k=1,\ldots K$ is a binary indicator for treatment assignment at the physician level. We assumed that time until osteoporosis management and censoring were conditionally independent given the physician cluster and that the patients in each treatment group shared a common survival distribution, hazard, and cumulative hazard function conditioned on the physician visited. We observed the minimum follow-up time T^o_{ki} between T_{ki} and D_{ki} and indicated if censoring occurred by C_{ki}, where $C_{ki}=I\{T^o_{ki}\le D_{ki}\}$ so that our data consisted of $(T^o_{ki}, C_{ki}; k=1,\ldots, K, i=1,\ldots,n_k)$.

A random effect w_k with density $f_\theta(.)$ is included in the model to measure physician to physician differences in treatment of patients at risk of osteoporosis. In the shared frailty model for the hazard,

$$\lambda_{ki}(t) = w_k \lambda_0(t) exp(\beta x_k).$$

$\lambda_{ki}(t); k=1,\ldots,K, i=1,\ldots,n_k$, predicts the patient specific hazard at time t given X_k and w_k. $\lambda_0(t)$ is the baseline hazard and β is the coefficient for the treatment effect [11]. If we were to combine w_k with the baseline hazard $\lambda_0(t)$, we could interpret the shared frailty model as a hazard model where the baseline hazard is shared among patients with the same physician.