Computational Statistics and Data Analysis 95 (2016) 150–160

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Comparing conditional survival functions with missing population marks in a competing risks model

Dipankar Bandyopadhyay [a,*], M. Amalia Jácome [b]

[a] Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA

[b] MODES Research Group, Faculty of Science, University of A Coruña, Spain

ARTICLE INFO

Article history:
Received 23 June 2014
Received in revised form 22 September 2015
Accepted 1 October 2015
Available online 19 October 2015

Keywords:
Competing risk
Fractional risk set
Logrank test
Right censoring
Weighted Kaplan–Meier

ABSTRACT

In studies involving nonparametric testing of the equality of two or more survival distributions, the survival curves can exhibit a wide variety of behaviors such as proportional hazards, early/late differences, and crossing hazards. As alternatives to the classical logrank test, the weighted Kaplan–Meier (WKM) type statistic and their variations were developed to handle these situations. However, their applicability is limited to cases where the population membership is available for all observations, including the right censored ones. Quite often, failure time data are confronted with missing population marks for the censored observations. To alleviate this, a new WKM-type test is introduced based on imputed population marks for the censored observations leading to fractional at-risk sets that estimate the underlying risk for the process. The asymptotic normality of the proposed test under the null hypothesis is established, and the finite sample properties in terms of empirical size and power are studied through a simulation study. Finally, the new test is applied on a study of subjects undergoing bone marrow transplantation.

1. Introduction

In medical and epidemiological follow-up studies, testing equality of two (or more) survival curves is a commonplace in understanding the effectiveness of treatment or exposure on the corresponding populations. Under the basic assumption of independent samples from two or more populations (including those whose failure times are right-censored), one resorts to the usual rank-based tests, which encompass the logrank (LR) (Mantel, 1966; Cox, 1972; Tarone and Ware, 1977), and the generalized Wilcoxon (Gehan, 1965; Breslow, 1970; Peto and Peto, 1972) statistics, or a Cox's regression (Klein and Moeschberger, 2003). In a typical competing risk framework where a subject in a healthy state may die due to one of J different risks, one might be interested in comparing the sub-populations (Bandyopadhyay and Datta, 2008) corresponding to the different death types. For example, targeted cancer therapies and vaccine trials devised to reduce the hazard of a specific disease mortality are not expected to affect the mortality from other causes or diseases, so the survival of the individuals dying from that specific disease is expected to be higher than those failing from other causes. However, the setup is complicated by the fact that the marks indicating the cause of death remain unavailable for subjects who were right censored, say, remaining alive at study completion. The cause of death can only be found (via. autopsy or otherwise) after the subject actually dies. In addition to the right censored subjects, determination of the cause of death might be prohibitive

* Correspondence to: Department of Biostatistics, Virginia Commonwealth University, 830 East Main Street, One Capitol Square, 7th Floor, PO Box 980032, Richmond, VA 23298-0032, USA. Tel.: +1 804 827 2058; fax: +1 804 828 8900.
E-mail address: dbandyop@vcu.edu (D. Bandyopadhyay).

http://dx.doi.org/10.1016/j.csda.2015.10.001
0167-9473/

for some subjects, thereby contributing to missing marks (Andersen et al., 1996). Removing those incomplete observations would maintain the correct size in the resulting test, however may lead to a loss of power (Bandyopadhyay and Datta, 2008).

Under a competing risk (or multistate) framework, there is a thick body of literature that focuses on estimation of certain (survival) quantities of interest, such as cause-specific hazard (CSH), cumulative incidence function (CIF), as well as regression functions. One of the basic assumptions here is that each subject is clearly identified to belong to one of the independent populations under consideration, including the right-censored ones. Hence, these functions may not be appropriate for summary probabilities of these competing causes. For example, in any clinical trial of a particular cancer type, the patients may die either out of the *cancer* or *other causes*, with a proportion experiencing censoring, say at study completion. The CIF can estimate the chance of a random patient dying due to the cancer before a pre-specified age, say 60. However, it does not answer the question: 'Among all patients dying due to the cancer (which also includes a proportion of the censored patients eventually experiencing cancer death), what is the probability that a random patient will die before age 60?'. Hence, the correct quantity to estimate and subject to hypothesis testing in this case is the *conditional survival function*, conditional on the cause of failure (Bandyopadhyay and Jácome, 2010). For this, the study subjects/individuals need to be split into J sub-populations, according to their eventual cause of failure. A more formal definition of this problem is available in Section 2.

Under missing population marks, a limited number of papers have dealt with this two-sample problem. Goetghebeur and Ryan (1990) derived a modified logrank test for a two-group comparison problem, while Dewanji (1992) provided a modification to that approach. Using a partial likelihood constructed under semiparametric assumptions, Goetghebeur and Ryan (1995) proposed a score test. Tsiatis et al. (2002) studied a combined logrank test assuming a logistic model for the conditional probability of dying out of the cause of interest and multiple imputation techniques (see e.g. Rubin, 1987) to impute missing causes of failure. The aforementioned approaches are based on several parametric assumptions, and therefore there is a clear risk of misspecification of the parametric model. More recently, Bandyopadhyay and Datta (2008) proposed a nonparametric weighted logrank (WLR)-type test for testing conditional survival functions, adapted to the missing population marks setup. Their approach is based on the fractional risks sets (FRS) proposition. The idea of FRS is to assign fractional probability masses to the censored observations (that has missing population marks) that represent the probability of belonging to each sub-population. That estimate is computed using the nonparametric maximum-likelihood estimator (NLMLE) of the transition probability (further details appear in Section 2). Bandyopadhyay and Datta (2008) showed that the performance of their WLR-FRS test is comparable to the classical WLR test when the population marks are known; however, for missing population marks, the WLR-FRS test outperforms the classical WLR test applied by throwing away the censored observations, while maintaining the size under the null.

The WLR test (both the classical and the FRS versions) is inarguably the most efficient test under the local alternative of proportional hazards (PH) among the two survival functions under comparison. However, it is not always sensitive to the stochastic ordering alternatives, particularly for crossing hazards. In practice, the assumption of a PH between the study and the control groups is unlikely to be true. For example, in surgical interventions, treated patients usually show less favorable short-term results versus placebo but far better long-term results; in the comparison of high/low doses in cancer chemotherapy, high dose may be ineffective initially, but produces favorable long-term results; or in cancer screening trials during a long follow-up period, variations can affect the shape of the observed differences between control and intervention groups (Shapiro et al., 1988). Considering these disadvantages of the WLR test, Pepe and Fleming (1989) introduced the weighted Kaplan–Meier (WKM) statistics as an alternative to the rank-based methods. The WKM test is based on the integrated weighted differences of the Kaplan–Meier (Kaplan and Meier, 1958) estimators of the corresponding survival curves. These WKM statistics are more sensitive to the magnitude of the survival differences than the rank-based tests, compare extremely well with the WLR test, and may perform far better than the WLR test under the crossing hazard alternatives. The asymptotic properties of the WKM test can be found in Pepe and Fleming (1991), and some extensions of the WKM test have been investigated by many authors, including Murray (2001), Shen and Cai (2001), Chi (2005), Lee et al. (2008) and (Lee, 2011), among others.

In order to alleviate the drawbacks of the WLR tests (both classical and FRS), this paper applies the FRS technique to the WKM test to accommodate missing causes of failure. Thus, the additional contribution is construction of a test where the WLR-FRS test of Bandyopadhyay and Datta (2008) becomes inefficient. The remainder of the paper is organized as follows. In Section 2, the new WKM-FRS test is introduced, and its asymptotic properties are studied. The finite sample performance of the test are explored and compared to the WLR-FRS test in terms of size and power for a variety of alternatives in Section 3. In Section 4, the proposed test is applied to a real dataset on patients undergoing bone marrow transplantation (BMT). Finally, some concluding comments are presented in Section 5 followed by an Appendix containing proofs of the theorems introduced in Section 2.

2. The WKM-FRS test

2.1. Background and notation

Consider a simple competing risks framework where a set of n subjects are exposed to $J = 2$ competing causes of failure/death, and let $L_j,\ j = 1, 2$ represent their (latent) failure times, that is, the potential survival times in hypothetical conditions where the only possible risk is the jth cause. Assuming no censoring, the observable random variables are (T, X),