Reliability Engineering and System Safety 125 (2014) 145–152

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

A nonparametric approach to medical survival data: Uncertainty in the context of risk in mortality analysis

Kateřina Janurová, Radim Briš *

VŠB—Technical University of Ostrava, Ostrava—Poruba, Czech Republic

ARTICLE INFO

Available online 15 April 2013

Keywords:
Survival analysis
Uncertainty
Medical data
Kaplan–Meier estimator
Nelson–Aalen estimator
Nonparametric predictive inference

ABSTRACT

Medical survival right-censored data of about 850 patients are evaluated to analyze the uncertainty related to the risk of mortality on one hand and compare two basic surgery techniques in the context of risk of mortality on the other hand. Colorectal data come from patients who underwent colectomy in the University Hospital of Ostrava. Two basic surgery operating techniques are used for the colectomy: either traditional (open) or minimally invasive (laparoscopic). Basic question arising at the colectomy operation is, which type of operation to choose to guarantee longer overall survival time.

Two non-parametric approaches have been used to quantify probability of mortality with uncertainties. In fact, complement of the probability to one, i.e. survival function with corresponding confidence levels is calculated and evaluated. First approach considers standard nonparametric estimators resulting from both the Kaplan–Meier estimator of survival function in connection with Greenwood's formula and the Nelson–Aalen estimator of cumulative hazard function including confidence interval for survival function as well. The second innovative approach, represented by Nonparametric Predictive Inference (NPI), uses lower and upper probabilities for quantifying uncertainty and provides a model of predictive survival function instead of the population survival function.

The traditional log-rank test on one hand and the nonparametric predictive comparison of two groups of lifetime data on the other hand have been compared to evaluate risk of mortality in the context of mentioned surgery techniques. The size of the difference between two groups of lifetime data has been considered and analyzed as well.

Both nonparametric approaches led to the same conclusion, that the minimally invasive operating technique guarantees the patient significantly longer survival time in comparison with the traditional operating technique.

1. Introduction

1.1. Motivation

Lifetime data in context with analysis of medical risk problems are usually denoted as survival or mortality data. The data are mostly applied to selected groups of patients trying to make a conclusion about differences between intended groups of patients. Survival data often contain right-censored observations arising from various situations. For example patient could still be alive after the last observation in the study, die due to other reasons than those of interest, move out or be withdrawn from the experiment. The study of colorectal surgery from the University Hospital of Ostrava, which this paper is engaged in, is not an exception. The retrospective observational study began in 2001 and has collected the data of about 850 patients. Its goal is to analyze the long-term survival data of cancer-diagnosed patients undergoing surgery of colon and rectum and decide whether the minimally invasive surgery technique guarantees longer overall survival time of the patients than the traditional open surgery technique. Patients were assigned to either of the two treatment groups on the basis of standard randomization mechanism used in given hospital.

The rate of minimally invasive techniques has been rising markedly since the 1990s in all branches of surgery and partly or even totally replaced traditional open techniques in some cases. The colorectal surgery is not an exception. Minimally invasive surgery has evident advantages: it is generally associated with lower operative stress and more favorable post-operative course. On the other hand there are many negative factors in using laparoscopic techniques in colorectal surgery, which can participate in morbidity and mortality in large measure (e.g. the risk of capnoperitoneum, longer operative time and extreme positioning of patients).

1.2. Bibliographic remarks

Various medical analyses compared the morbidity and mortality after both types of surgeries, for example the conclusion of

* Corresponding author. Tel.: +420 59 732 3334.
E-mail address: radim.bris@vsb.cz (R. Briš).

0951-8320/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ress.2013.03.014

European Association of Endoscopic Surgery for colon carcinoma states, that there is no difference between morbidity and mortality of laparoscopic and open operations of colon [1]. A lot of information regarding the laparoscopic colectomy for cancer is now available from many clinical studies [2], [3–7], which proved merits of the laparoscopic surgery.

Much less information is nowadays available regarding the cancer of rectum. Meta-analyses comparing laparoscopic versus open surgery for rectal cancer are very rare and in fact they are mostly connected with short term results. Therefore laparoscopic surgery for rectal cancer is still an open problem in recent time and especially analyses of long term outcomes are eagerly awaited.

The standard nonparametric approach for comparing two groups of survival data includes the estimation of survival function. Kaplan and Meier derived their estimator of survival function in [8], where they also showed that it could be interpreted as a nonparametric maximum likelihood estimator and derived proper expression for its variance. The large sample properties of the Kaplan–Meier estimator were studied by Efron [9] and later by Breslow and Crowley [10].

The Nelson estimator of cumulative hazard function was originally proposed as a graphical method for checking exponentiality [11], because the cumulative hazard function is linear when the underlying failure time distribution is exponential. Aalen suggested this estimator in the multiplicative intensity model [12] and his work has become a fundamental source for the analysis of survival data based on counting process approach. Andersen, Borgan, Gill and Keiding [13] and Fleming and Harrington [14] discussed precisely the statistical theory and use of the Nelson–Aalen estimator. Further discussion is provided for example by Johansen [15], Jacobsen [16], Karr [17] and Gill [18].

Mantel derived the hypothesis test for comparing survival distributions of two groups in [19] that is known as the log-rank test due to Peto and Peto [20].

The innovative nonparametric approach, Nonparametric Predictive Inference (NPI), has been developed mainly by Coolen and his colleagues since the middle 1990s [21]. NPI is a statistical method that depends on few modeling assumptions and quantifies uncertainty by the use of interval-valued probability (lower and upper probability) [22,23]. NPI is based on the Hill assumption $A_{(n)}$ [24] which predicts the underlying distribution under the assumption of finite exchangeability of random quantities [25]. Berliner and Hill [26] generalized $A_{(n)}$ for survival analysis with right-censored observations. Related concept was developed further by Coolen and Yan, who generalized Hill's $A_{(n)}$ in 'right-censoring $A_{(n)}$' to take the exact censoring information into account [27]. They also dealt with the nonparametric predictive comparison of two groups of lifetime data [28] and with multiple comparisons Maturi [29]. The size of the difference between groups was also considered by Coolen and Al-Nefaiee [30]. In [31] the NPI methods for accuracy of diagnostic tests with continuous test results are presented and discussed. The paper demonstrates how NPI can be used for data that did not contain censored observations to compare two continuous diagnostic tests.

There are also many other applications of NPI including solution of the problems in reliability, statistics and operational research.

1.3. Main contributions of the paper

In this paper we discuss the nonparametric approach for right-censored medical data assuming throughout that the censoring process is independent of the mortality process.

In analyzing uncertainty due to sampling variation in medical survival data, the results of three nonparametric methods have been compared. The traditional nonparametric Kaplan–Meier survivor function and Nelson–Aalen cumulative hazard estimators are contrasted with nonparametric predictive inference (NPI), an innovative recently proposed alternative based on the conceptually novel notion of imprecise partially specified interval probabilities from which minimum and maximum bounds on survival probabilities can be derived.

The paper has a dual purpose. Firstly to demonstrate how techniques of nonparametric analysis can be used to determine whether possible differences in mortality risk between surgical procedures can be reliably assessed in the face of uncertainty due to sampling variation. Secondly to use this analysis as a platform to identify characteristics of these conceptually distinct analytic approaches that may lead to conflicting, complimentary, or augmentative results. The basis for comparison will be traditional and NPI type estimates of survival probabilities, log rank significance tests, and NPI derived stochastic dominance bounds.

2. Standard non-parametric approach

When we apply survival analysis to medical data, we are generally interested in describing the life expectancy of objects under study, rather than the mortality rate. That is the reason, why we concentrate in medical investigations on estimating survival functions. It represents the probability that the survival time of patient is greater than some specified time t, denoted $S(t)=P(T>t)$. In this section, long term survival analysis using the traditional nonparametric Kaplan–Meier and Nelson–Aalen estimators is reviewed.

2.1. Survival data

Let us assume lifetime survival data with some values randomly censored independently of survival time. Our experiment generates n independent observations in the following form:

$$(t_1, c_1), \ldots, (t_n, c_n),$$

where t_i is either a time of death or a time in which the observation of i-th patient is stopped (withdrawn); and $c_i=1$ (resp. $c_i=0$) is censoring indicator, according to death (resp. stopping time) occurring first.

We assume that of the n experimental observations there are m observed deaths ($m \leq n$) and $n-m$ withdrawn or censored observations with no ties. The sample is then rank ordered $t_{(1)} < \ldots < t_{(n)}$. Let $c_{(i)}$ be the indicator corresponding to $t_{(i)}$, $i=1,\ldots,n$.

2.2. Kaplan–Meier estimator of the survival function

The Kaplan–Meier estimator of the survival function, $S(t)$ represents at each observed survival time t the fraction of patients still alive after the treatment and is computed from the product of estimates of conditional survival probability between successive observed times of death in the ordered sample.

$$\widehat{S}(t) = \prod_{t_{(i)} \leq t, c_{(i)}=1} \frac{n_i - d_i}{n_i}, \tag{1}$$

where n_i = number of patients at risk of dying until $t_{(i)}$ (the time $t_{(i)}$ is not included); d_i = number of observed deaths. The asymptotic normality of the Kaplan–Meier estimator and functions of it can be established using counting process theory [13,14].