Statistics and Probability Letters 109 (2016) 152–158

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Nonparametric restricted mean analysis across multiple follow-up intervals

Nabihah Tayob[a,*], Susan Murray[b]

[a] The University of Texas MD Anderson Cancer Center, Department of Biostatistics, 1400 Pressler St, Houston, TX 77030, USA

[b] University of Michigan, Department of Biostatistics, Ann Arbor, MI 48109, USA

ARTICLE INFO

Article history:
Received 21 April 2015
Received in revised form 11 November 2015
Accepted 12 November 2015
Available online 27 November 2015

Keywords:
Correlated times-to-events
Follow-up intervals
Residual life
Restricted mean survival

ABSTRACT

We provide a nonparametric estimate of τ-restricted mean survival using follow-up information beyond τ when appropriate to improve precision. The variance accounts for correlation between follow-up windows. Both asymptotic calculations and simulation studies recommend follow-up intervals spaced approximately $\tau/2$ apart.

1. Introduction

Yearly progression predictions are commonly reported for clinical longitudinal data. For example, Raghu et al. (2011) report that mild to moderate idiopathic pulmonary fibrosis (IPF) patients tend to lose 0.2 l in forced vital capacity lung function per year. This is a valuable summary statistic that has not been sufficiently explored for censored time to event data. It would be clinically useful to be able to report that IPF patients followed for 10 years were observed to live 91% of each year, on average, given they were alive at the start of the year. This concise estimate describes how the disease will affect patients in the short term and indicates stability of the disease where appropriate. Health economists have noted that patient's value life-years closer to the present more than those in the future (Gyrd-Hansen and Sogaard, 1998).

The restricted mean residual life function (RMRL) is the expected days of life per year for those surviving at the beginning of the year and may be used to view trends in these summary statistics over time. Ghorai et al. (1982) proposed an estimator based on integrated conditional Kaplan–Meier estimates (Ghorai and Rejto, 1987) and Na and Kim (1999) proposed a smooth-spline estimator for this quantity, among others. Yu (2003) developed confidence bands for restricted mean residual life functions estimated via Nelson–Aalen estimates, Cox model hazard estimates and inverse weighted hazard estimates, calling them expected life prosper functions (ELPF). Stability of these functions suggests an opportunity for producing an overall summary statistic that is more precise.

In Section 2 we review RMRL estimation and confidence band construction. Section 3 describes the nonparametric τ-restricted mean survival estimator that combines information across different τ-length intervals of follow-up. The proposed variance described in Section 3 is based on a linearization of random components of the estimator, similar to the approach recommended by Woodruff (1971) and more recently Williams (1995). In Section 4 we consider how to choose the number

* Corresponding author. Tel.: +1 734 239 4235.
E-mail address: ntayob@mdanderson.org (N. Tayob).

http://dx.doi.org/10.1016/j.spl.2015.11.014
0167-7152/

of follow-up intervals useful for obtaining efficiency gains. A simulation study that assesses the performance of the proposed estimate and its variance against currently available competitors is presented in Section 5. Examples of the proposed analysis approach pertaining to IPF patients and diabetic retinopathy patients are given in Section 6. A discussion follows in Section 7.

2. τ-restricted mean residual life

2.1. Notation

For each of n patients we define observed event time, $X_i = \min(T_i, C_i)$, with failure indicator $\delta_i = I(T_i \leq C_i)$, based on true failure time, T_i, and censoring time, C_i, $i = 1, \ldots, n$. Calendar time, t, is measured from the start of the study. We define the residual life observed at t as $X_i(t) = (X_i - t)I(X_i \geq t)$ with failure indicator variable $\delta_i(t) = \delta_i I(X_i \geq t)$.

For a τ-length interval starting at calendar time, t, the τ-restricted mean residual lifetime is $\mu(t, \tau) = E\{\min(T - t, \tau)|T > t\} = \int_0^\tau Pr(T > t + u|T > t)du$. Here, u denotes an internal time scale measured from calendar time t. We examine values of $\mu(t, \tau)$ measured at different calendar times $t \in \{t_1, \ldots, t_b\}$, with $t_1 = 0$ in all that follows.

Counting process notation includes the two timescales described, t and u. For individual i, the event counting process is $N_i(t, u) = I\{X_i(t) \leq u, \delta_i(t) = 1\}$ with at-risk process $Y_i(t, u) = I\{X_i(t) \geq u\}$. At t, the total number of events occurring no later than u is $N(t, u) = \sum_{i=1}^n N_i(t, u)$ and the number at risk at u is $Y(t, u) = \sum_{i=1}^n Y_i(t, u)$. We require notation combining counting process quantities across calendar times, $\{t_1, \ldots, t_b\}$. For individual i and internal time u, $N_i(u) = \sum_{k=1}^b N_i(t_k, u)$ and $Y_i(u) = \sum_{k=1}^b Y_i(t_k, u)$. Combining information across follow-up intervals and patients gives us $N(u) = \sum_{k=1}^b N(t_k, u)$, total number of events occurring no later than u, and $Y(u) = \sum_{k=1}^b Y(t_k, u)$, total number at risk at u.

2.2. Estimation of τ-restricted mean residual life

The τ-restricted mean residual lifetime function, $\mu(t, \tau)$, tracks subsequent expected lifetime during an interval of length τ given the patient has survived up to time t. Henceforth we call $\mu(t, \tau)$ the RMRL function. Using notation from the previous section, a consistent nonparametric estimator of the RMRL function is $\hat{\mu}(t, \tau) = \int_0^\tau \hat{P}(t, s)ds$ where $\hat{P}(t, s) = \exp\left\{-\int_0^s dN(t, u)/Y(t, u)\right\}$.

2.3. Confidence bands

We slightly modify work from Yu (2003) for RMRL confidence band calculations applied to times $\{t_1, \ldots, t_b\}$. Plotting the RMRL function with its corresponding bands is useful in suggesting whether follow-up windows may be combined for estimation or not.

Technical development of confidence bands for $\{\mu(t_1, \tau), \ldots, \mu(t_b, \tau)\}$ is based on the Gaussian process $B(t, \tau) = n^{1/2}\{\hat{\mu}(t, \tau) - \mu(t, \tau)\}$ and the covariance of this process at times t_{j_1} and t_{j_2}, $j_1 = 1, \ldots, b$, $j_2 = 1, \ldots, b$. The estimated covariance matrix $\hat{\sum}_{b\times b}$, with estimates of $\mathrm{cov}\{B(t_{j_1}, \tau), B(t_{j_2}, \tau)\}$ as the (j_1, j_2) elements, is

$$n\int_0^\tau \int_0^\tau \hat{P}(t_{j_1}, s)\hat{P}(t_{j_2}, s')\sum_{i=1}^n \int_0^s \int_0^{s'} \frac{dN_i(t_{j_1}, u)dN_i(t_{j_2}, v)}{Y(t_{j_1}, u)Y(t_{j_2}, v)}ds\, ds'.$$

Following Lin et al. (1994), the asymptotic distribution of $\{B(t_1, \tau), \ldots, B(t_b, \tau)\}$ is approximated by generating a large number of mean zero multivariate Normal samples using the observed covariance structure, $\hat{\sum}_{b\times b}$. Using these samples, we calculate q_α satisfying $Pr\{\max_{t\in\{t_1,\ldots,t_b\}} |\hat{B}(t, \tau)| > q_\alpha\} = \alpha$. Level $100(1-\alpha)$% confidence band values surrounding $\mu(t, \tau)$ become $\left[\hat{\mu}(t, \tau) - n^{-1/2} \times q_\alpha, \hat{\mu}(t, \tau) + n^{-1/2} \times q_\alpha\right]$, calculated at times $t = \{t_1, \ldots, t_b\}$. In practice, these confidence bands perform well provided that there are at least 25 event times following t_b (Yu, 2003).

3. Overall τ-restricted mean survival

The RMRL plots and associated confidence bands are a useful diagnostic tool to assist in deciding whether disease progression is stable, i.e. the RMRL is the same at each $t_1, \ldots, t_b$, or not. When the RMRL and its corresponding confidence bands do not indicate a strong trend, we develop a more efficient estimate of the expected number of days lived in the next τ years by pooling appropriate follow-up periods beginning at times $t \in \{t_1, \ldots, t_b\}$.

3.1. Estimation

The proposed estimate of the overall τ-restricted mean survival is

$$\hat{\mu}^*(\tau) = \int_0^\tau \exp\left\{-\int_0^s \frac{dN(u)}{Y(u)}\right\} ds. \quad (1)$$