Journal of Statistical Planning and Inference 182 (2017) 1–11

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

# On the conditional probability for assessing time dependence of association in shared frailty models with bivariate current status data

Steffen Unkel

*Department of Medical Statistics, University Medical Center Göttingen, Humboldtallee 32, 37073 Göttingen, Germany*

ARTICLE INFO

*Article history:*
Received 2 March 2016
Received in revised form 5 May 2016
Accepted 1 September 2016
Available online 20 October 2016

*Keywords:*
Conditional probability
Current status data
Cross-sectional data
Frailty models
Heterogeneity
Time-varying association

ABSTRACT

Shared frailty models are frequently used for inducing dependence between survival times. In this paper, we consider bivariate current status data that are reasonable to model by shared frailty models. A time-dependent association measure that has a conditional probability interpretation is revisited for its potential application to such data. We propose a method of estimation and derive asymptotic standard errors for this measure. Its small sample performance and its performance in assessing the temporal variation in the strength of association in realistic scenarios is investigated by means of experiments. We show that the measure based on the conditional probability can vary with time even in the absence of any time-dependent effects. Furthermore, we give evidence that it lacks interpretability in suggesting appropriate frailty models. We provide an illustration with multivariate current status data arising from a community-based study of cardiovascular diseases in Taiwan. We compare the observed patterns of association with the ones obtained by employing a fairly new time-varying association measure that is relevant for shared frailty models, owing to its connection to the cross-ratio function, and which serves as a diagnostic tool for suggesting classes of frailty distributions with constant, increasing or decreasing association over time.



## 1. Introduction

Bivariate current status data can be formally represented as $\{X, \delta_1 = I(T_1 \leq X), \delta_2 = I(T_2 \leq X)\}$, where $I$ denotes the indicator function, $T_1$ and $T_2$ are the failure times of interest and $X$ is the monitoring time at which $T_1$ and $T_2$ are measured from the same observational units and that is independent of the failure times (Jewell et al., 2005; Sun, 2006). In this paper, we consider bivariate current status data that are reasonable to model by shared frailty models (Duchateau and Janssen, 2008; Hougaard, 2000; Wienke, 2011), with the frailty solely generating the association structure between $T_1$ and $T_2$ and the variability between the observational units, also referred to as the heterogeneity in the data, being represented by the variance of the frailty. Such bivariate current status data arise in various fields (Jewell et al., 2005). Consider for example tumorigenicity experiments on a single non-lethal tumor at two different sites, e.g. liver and brain, to investigate whether the environment accelerates the time until tumor onset in animals. In these experiments, the time to tumor onset in the animals is only known to be less than or greater than the observed time of death or sacrifice. A shared frailty model is natural in this setting, the shared latent frailty representing environmental exposures relevant to the progression of disease at different sites.

*E-mail address:* steffen.unkel@med.uni-goettingen.de.

http://dx.doi.org/10.1016/j.jspi.2016.09.009
0378-3758/

Another example arises in twin pair studies in genetics, where the phenotypes of interest are the ages at onset of a specific disease. For neurological disorders such as Alzheimer's disease, the exact age of onset is usually not known even when a definitive diagnosis is available. If $T_j (j = 1, 2)$ is the unknown age of onset for the $j$th twin, then in such cases only bivariate current status information $(\delta_1, \delta_2)$ is observed instead of $(T_1, T_2)$. In other words, it is only known at the monitoring time whether the $j$th twin has the disease or not. Interest may focus on the strength of association between $T_1$ and $T_2$ for both monozygotic and dizygotic twins. Again, a shared frailty model is natural, the latent frailty variable representing genetic characteristics that may have a bearing on the onset of the disease of interest.

In some circumstances, it is also of interest to assess the time dependence of association (Anderson et al., 1992; Oakes, 1989), for example to investigate the age-varying influence of genetic factors on the disease-free life expectancy of individuals by comparing the disease-free life spans of monozygotic and dizygotic twins.

In shared frailty models for bivariate survival data, the frailty distribution is identifiable through Clayton's local cross-ratio function (Clayton, 1978), which describes how the heterogeneity of the hazard functions in the survivor population evolves over time. Hence, the cross-ratio function may serve as a diagnostic tool in an exploratory analysis for suggesting appropriate frailty distributions and assessing the temporal variation in the strength of association in bivariate survival data (Farrington et al., 2012; Viswanathan and Manatunga, 2001). This association measure is unavailable for current status data, though.

The odds ratio is the most obvious and popular association measure for binary data and is widely used in many fields, such as in epidemiological studies as a measure of association between the occurrence of a particular disease state or condition and an exposure factor (Jewell, 2003). It can easily be estimated after fitting a linear logistic model to dichotomous data (Collett, 2002). The odds ratio can be computed from current status data. However, the odds ratio suffers the disadvantages that it can vary with time even in the absence of any time-dependent effects and that it does not reliably suggest appropriate frailty distributions (Unkel and Farrington, 2012).

Anderson et al. (1992) introduced the following time-dependent measure for association based on the conditional probability:

$$\psi(t_1, t_2) = \frac{\mathrm{P}(T_1 > t_1 | T_2 > t_2)}{\mathrm{P}(T_1 > t_1)} = \frac{S(t_1, t_2)}{S_1(t_1) S_2(t_2)}, \tag{1}$$

where $S_j(t_j) = \mathrm{P}(T_j > t_j)$ denotes the marginal survivor function for $T_j (j = 1, 2)$ and $S(t_1, t_2) = \mathrm{P}(T_1 > t_1, T_2 > t_2)$ is the joint survivor function. Large values of $\psi(t_1, t_2)$ indicate positive dependence between $T_1$ and $T_2$. For independent events $T_1 > t_1$ and $T_2 > t_2$, $\psi(t_1, t_2) = 1$. If $S(t_1, t_2) < S(t_1) S(t_2)$, then there is negative dependence between $T_1$ and $T_2$. In Anderson et al. (1992), the measure (1), indexed by age, is applied to right-censored data from the Danish Twin Registry to describe the differences in the strength of association between monozygotic and dizygotic twins with respect to their life spans and to investigate how these associations depend on the age of the twins.

Unfortunately, for current status data, the joint survivor function $S(t_1, t_2)$ is unobservable; only $S(x, x)$, where $X = x$ denotes the observed monitoring (censoring) time, is available along with the marginals $S_1(x) = S(x, 0)$ and $S_2(x) = S(0, x)$. This means that for current status data, one can assess the association between two survival variables by means of

$$\psi(x) = \frac{S(x, x)}{S_1(x) S_2(x)}. \tag{2}$$

In this paper, we investigate the measure (2) and pay particular attention to the evaluation of its usefulness to govern the temporal variation in the strength of association inherent in bivariate current status data and in serving as an exploratory tool for suggesting frailty distributions. We propose a method of estimation and introduce asymptotic standard errors for (2). We also apply an association measure introduced in Unkel and Farrington (2012), which is based on the association parameter derived from Clayton's copula (Clayton, 1978) for quantifying time-dependent association. This measure tracks the variation of the cross-ratio function with time. Therefore, it serves as a diagnostic tool for suggesting classes of frailty distributions with constant increasing or decreasing association over time. The shape of the observed time-varying association aids identification of a suitable frailty model, which then could be fitted to the data set at hand. Up to the author's knowledge, the association measure by Unkel and Farrington (2012) has been only applied so far to bivariate serological survey data on pairs of infections with similar and different transmission routes, where the time-varying association is likely to represent heterogeneities in activity levels and/or susceptibility to infection (see also Unkel et al., 2014; Farrington et al., 2013).

In the present paper, our methods are illustrated with multivariate current status data arising from a community-based study on three cardiovascular diseases in Taiwan. These data were originally analyzed by Wang and Ding (2000) under the assumption of constant pairwise association over time. We explore the possible time dependence of association in the data. The remainder of the paper is organized as follows. In Section 2, we present maximum likelihood estimators for the association measure based on the conditional probability and derive asymptotic standard errors. An evaluation of how the conditional probability measure performs with respect to identifying time-varying effects in shared frailty models with bivariate current status data is given in Section 3. We also investigate the finite sample performance of the association measure in realistic scenarios using simulations. In Section 4, the methodology discussed in this paper is applied to the aforementioned data set. Concluding comments are given in Section 5. Computations for this paper were carried out using the software package R version 3.2.1 (R Core Team, 2015). All computer code used is available upon request.