

Pseudo-likelihood estimation for a marginal multivariate survival model

Fabián Tibaldi^{*,†}, Geert Molenberghs, Tomasz Burzykowski and Helena Geys

*Center for Statistics, Limburgs Universitair Centrum, Transnationale Universiteit Limburg, Diepenbeek
Universitaire Campus, B3590 Diepenbeek, Belgium*

SUMMARY

In this paper, we propose a multivariate Plackett–Dale model for survival outcomes. A pseudo-likelihood method for the estimation of the parameters is proposed and these ideas are applied to two case studies. The modelling approach is similar in spirit but different from Parner’s approach. The first study is in AIDS, where the overall survival time and different opportunistic infections in HIV-infected patients are studied. The second study is on adoption data where the association of the survival times within families is modelled, illustrating the use of the proposed methodology for the context of population genetics. Copyright © 2004 John Wiley & Sons, Ltd.

1. INTRODUCTION

Survival models have been used intensively during the past two decades, across a number of application areas. Medical researchers used them extensively, but in many other fields where the main interest is in time-to-event, they became an important tool as well [1]. The effect of one or more covariates on the patient’s survival can be modelled via the Cox model [2], but we should recall that independence of survival times from one observation to the other is one of the basic assumptions of this model. However, in the last few years there has been an increasing interest in frameworks where two or more events per patient or per statistical unit are observed. These statistical units can refer to clusters and hence multivariate survival models should be used, taking into account within-cluster dependencies. The former phenomenon is observed in groups of patients that share common characteristics, such as in family studies where the members share genetic and environmental factors. There are several issues that one should take into account when extending the Cox model or other univariate survival model, to the situation where the association needs to be modelled, which is the topic of the current paper. The key idea is to introduce a model that allows for a full association structure between the times to event pertaining to a given unit while, due to an appropriate use of pseudo-likelihood ideas, keeping the computational burden under control.

*Correspondence to: Fabián Tibaldi, Center for Statistics, Limburgs Universitair Centrum, Universiteit Limburg, Universitaire Campus, B3590 Diepenbeek, Belgium.

†E-mail: fabian.tibaldi@luc.ac.be

The paper is organized as follows. Section 2 motivates the problem through two case studies. Section 3.1 gives a description of the Plackett–Dale model [3] for survival data in the bivariate case. Section 3.2 describes an extension of the model to the case of k correlated survival times and proposes a pseudo-likelihood approach for the estimation of the parameters of the model. Section 5 contains the analysis of the case studies.

2. MOTIVATING CASES

In this section, we introduce two different studies for which our methodology is of use. The AIDS case study deals with intrasubject correlation, i.e. multiple events per subjects are recorded. The adoption study is an example of a study where clustering, within-cluster dependencies are present.

2.1. *The AIDS study*

These data arise from a randomized clinical trial. A total of 1530 patients who participated in two clinical trials sponsored by the AIDS Clinical Trials Group (ACTG): ACTG 116A [4] and 116B/117 [5] were randomized to compare zidovudine (ZDV) and two doses of didanosine (ddI). Participants either had a diagnosis of AIDS or AIDS-related complex (ARC) and/or had CD4 counts of 300 or fewer. The primary outcomes of interest for this analysis were survival and new or recurrent AIDS-defining events. Patients were randomly assigned to receive one of the following three treatments: ddI 750 mg per day, ddI 500 mg per day or ZDV 600 mg per day. These studies enrolled patients between October 1989 and April 1991; patients were followed for a median of 65 weeks and a maximum of 132 weeks. For illustration, ZDV is compared to any dose of DDI; therefore we use a binary indicator variable for treatment effect. Measures of CD4 for individual patients are included in the model. This choice is supported by the work of Saah *et al.* [6], who found that CD4 was a laboratory measure in a Cox proportional hazards model which predicted survival after AIDS. There has been some debate in the literature as to whether a dichotomization of CD4 can be justified or not. We will use a continuous version of this variable but any other categorization can be considered without substantially having to modify the methodology. Molenberghs *et al.* [7] studied the joint modelling of survival and CD4 count on these data.

2.2. *The adoption study*

This study presented in Reference [8] was carried out to analyse the impact of environmental and genetic factors on the survival of adult adoptees. To this end, dependencies between the survival time of children and biological parents, and between children and adoptive parents are the focus of interest. In this study, families with adoptive children, born between 1924 and 1926, were analysed. The basic idea is that association between survival times of biological parents and children can be assigned to some extent to genetic factors, while associations between children and adoptive parents can be due only to environmental factors.

These data were studied by Nielsen *et al.* [9], who proposed a shared gamma frailty model and by Parner [10], who proposed a composite likelihood method for the estimation of the frailty parameters and the standard deviations. We propose to use a Plackett–Dale model [11] for correlated survival time data with Weibull margins, as described in the next section.

3. MODEL DESCRIPTION

3.1. Bivariate Plackett–Dale model for survival data

In this section, we introduce the Plackett–Dale model for two survival outcomes. Assume that T_1 and T_2 are correlated survival times, then the joint survival function of (T_1, T_2) can be written as

$$S_{T_1 T_2}(t_1, t_2) = P(T_1 \geq t_1, T_2 \geq t_2) = C_{\theta_{12}}\{S_{T_1}(t_1), S_{T_2}(t_2)\}, \quad t_1, t_2 \geq 0 \tag{1}$$

where S_{T_1} and S_{T_2} denote marginal survival functions and $C_{\theta_{12}}$ is a copula. An attractive feature of model (1) is that the margins do not depend on the choice of the copula function.

In principle, in model (1) any copula function can be used. For simplicity, we consider primarily one-parameter families; hence the use of a single parameter θ_{12} in (1). Some possible options are the Clayton, Hougaard and Plackett copulas. Burzykowski *et al.* [11] studied them in detail within the framework of surrogate endpoints. For the Clayton and Hougaard copulas, model (1) reduces to a proportional frailty model [12] with frailties generated, respectively, by the gamma and the positive stable distributions.

To model the effect of specific covariates on the marginal distributions of T_1 and T_2 in (1) we propose to use the proportional hazard model:

$$S_{T_k}(t_k) = \exp \left\{ - \int_0^{t_k} h_{T_k}(x) \exp(\boldsymbol{\beta}_{T_k} \mathbf{Z}_k) dx \right\}, \quad k = 1, 2 \tag{2}$$

where h_{T_1} and h_{T_2} are marginal baseline hazard functions and $\boldsymbol{\beta}_{T_1}$ and $\boldsymbol{\beta}_{T_2}$ are vectors of unknown regression parameters corresponding to the covariates \mathbf{Z} . The hazard functions can be specified parametrically or can be left unspecified as in the classical model proposed by Cox [3]. When the hazard functions are specified, maximum likelihood estimates of the parameters for joint model (1) and (2) can be obtained [13]. Alternatively, the two-stage parametric procedure proposed by Shih and Louis [14] can be used, in which parameters of the marginal survival functions S_{T_1} and S_{T_2} are estimated first (assuming independence), and then θ_{12} is estimated conditional on the estimated values of the marginal parameters.

This one-parameter family is closely related to the Plackett family of bivariate distributions [15]. In this case the dependence can be defined using a *global cross-ratio* at (t_1, t_2) which, given the marginal cumulative density functions F_{T_1} and F_{T_2} , is given by

$$\theta_{12}(t_1, t_2) = \frac{F(t_1, t_2)[1 - F_{T_1}(t_1) - F_{T_2}(t_2) + F(t_1, t_2)]}{[F_{T_1}(t_1) - F(t_1, t_2)][F_{T_2}(t_2) - F(t_1, t_2)]} \tag{3}$$

Note that ‘global’ refers to the fact that, at every point, the bivariate space is divided into four quadrants. Then, the probability over each quadrant is calculated and these four quantities are then used to compute the odds ratio. Here, $\theta_{12} = \theta_{12}(t_1, t_2)$ satisfies $0 \leq \theta_{12} \leq \infty$ when $F(t_1, t_2)$ satisfies the Fréchet–Hoeffding [16] bounds. The components in (3) are the quadrant probabilities in \mathbb{R}^2 with vertex at (t_1, t_2) . The Plackett distribution is obtained for constant cross-ratio $\theta_{12}(t_1, t_2) \equiv \theta$ [15, 17]. The joint distribution $F_{T_1 T_2}$ is defined by means of (3), when F_{T_1} , F_{T_2} and θ_{12} are known.

The values of the Plackett distributions are found as one of the two solutions of the following second degree polynomial equation if the marginal distribution functions F_{T_1} and F_{T_2} ,

and the cross-ratio θ_{12} are known:

$$\theta_{12}(F - F_{T_1})(F - F_{T_2}) - F[F - (F_{T_1} + F_{T_2} - 1)] = 0 \quad (4)$$

Dale [18] and Mardia [17] gave an explicit solution for (4):

$$F_{T_1, T_2}(t_1, t_2) = \begin{cases} \frac{1 + (F_{T_2}(t_2) + F_{T_1}(t_1))(\theta_{12} - 1) - H(F_{T_2}(t_2), F_{T_1}(t_1), \theta_{12})}{2(\theta_{12} - 1)} & \text{if } \theta_{12} \neq 1 \\ F_{T_2}(t_2)F_{T_1}(t_1) & \text{if } \theta_{12} = 1 \end{cases} \quad (5)$$

where

$$H(F_{T_1}, F_{T_2}, \theta_{12}) = \sqrt{(1 + (\theta_{12} - 1)(F_{T_1}(t_1) + F_{T_2}(t_2)))^2 + 4\theta_{12}(1 - \theta_{12})F_{T_1}(t_1)F_{T_2}(t_2)} \quad (6)$$

Note that the other solution to the polynomial can be shown to always lie outside the Fréchet bounds. Mardia [17] showed that $F_{T_1, T_2}(t_1, t_2)$ is always a bivariate copula, with θ_{12} in $[0, +\infty]$. Although (5) and (6) was obtained based on the defining equation for the distribution function F , it can be shown that exactly the same copula is obtained for the survival function $S = 1 - F$.

Based upon this distribution function, we can derive a bivariate Plackett *density* function $f_{T_1, T_2}(t_1, t_2)$ for two survival times using (5) and (6) by calculating $\partial F_{T_1, T_2}(t_1, t_2) / \partial t_1 \partial t_2$ in an appropriate way by taking into account censoring.

The parameters of this model and their standard deviations can be estimated by means of the maximum likelihood method. Appendix A details the expression for the log likelihood function, together with the derivatives of the distribution function F .

3.2. Multivariate Plackett–Dale model for survival data and pseudo-likelihood estimation

While the model described in Section 3.1 suffices to analyse bivariate time-to-event outcomes, an extension is needed for applications with more than two times. To this end, we consider an experiment involving N subjects or clusters of k time-to-event measurements.

The principal idea can be laid out in three steps. First, we construct a model for these k times by considering univariate models for every time-to-event separately. It is evident that covariates can be included in these parametric marginal models. Second, we consider bivariate models for every possible pair that can be formed from the k times and of which the univariate marginal models are the ones already considered; in other words, Plackett–Dale models will be considered for every possible pairs. Third, in order to avoid the full multivariate specification of the model, while nevertheless properly accounting for the full association structure, pseudo-likelihood ideas are used (similar to but different from the principles behind generalized estimating equations) to obtain valid point estimates as well as valid precision estimates.

This approach is similar in spirit to the one proposed by Parner [10] in the sense that both are marginal models for multivariate survival data and both use pseudo-likelihood-related ideas. However, the actual copulas chosen are different, enabling a comparison of the results from both, for example. Since there is no unambiguous choice as to what the best model would be for multivariate survival data, a more ample choice of models is desirable and can lead up to a sensitivity analysis.

Suppose that we also observe a vector of covariates \mathbf{Z} . A Weibull distribution is assumed for each time T_j with λ_{T_j} and p_{T_j} the scale and shape parameters, respectively. While we focus on Weibull marginals, different researchers may choose to use different univariate marginal survival distributions, implying only relatively small adaptations of the methodology. The information concerning subject i can be expressed in vector format as $(T_{i1}, \dots, T_{ik}, \Delta_{i1}, \dots, \Delta_{ik}, z_{i1}, \dots, z_{in_k})$, with n_k being the number of covariates, so that $\mathbf{W}_{ij} = (T_{ij}, \Delta_{ij}, \mathbf{Z}_i)$ are the values for a particular subject i and time point j .

While a full multivariate formulation of the Plackett–Dale model has been carried out in the context of ordinal data [3, 19], it poses non-trivial computational complexities. Instead, marginal pseudo-likelihood ideas will be used to keep the amount of computation under control, while enabling to answer relevant research questions [20–22].

The idea behind our pseudo-likelihood function is based on considering all possible pairs $(\mathbf{W}_{ir}, \mathbf{W}_{il})$ of outcomes on an individual, producing $f_{T_r T_l}(\mathbf{W}_{ir}, \mathbf{W}_{il})$, rather than the full multivariate density, and then taking the product over them. The resulting function will be denoted by PL and its log by

$$\ln p\ell(\Phi) = \sum_{i=1}^N p\ell_i \tag{7}$$

with

$$p\ell_i = \sum_{(s,t) \in S} \ln f_{T_s T_t}(\mathbf{W}_{is}, \mathbf{W}_{it}, \Phi)$$

where S is the set of indices with all possible pairs of outcomes of interest, $f_{T_s T_t}$ is the value of the function defined in Section 3.1 evaluated in the respective outcomes for subject i and Φ is the vector of parameters. Specifically $\Phi' = (\theta', \beta_T', \lambda_T', \mathbf{p}_T')$, with θ being the subvector of association parameters, β_T the subvector of coefficients corresponding to the covariates \mathbf{z} and λ_T and \mathbf{p}_T subvector of parameters from the Weibull distribution.

The pseudo-likelihood estimator $\hat{\Phi}$ is defined as the maximizer of (7). Consistency has been shown by Arnold and Strauss [23], Le Cessie and Van Houwelingen [20], and Geys, Molenberghs, and Ryan [22]. Precisely, it converges in probability to Φ_0 , the true parameter value and $\sqrt{N}(\hat{\Phi} - \Phi_0)$ converges in distribution to

$$N_q(\mathbf{0}, J(\Phi_0)^{-1} K(\Phi_0) J(\Phi_0)^{-1}) \tag{8}$$

with $J(\Phi)$ defined by

$$J_{rl} = - \sum_{(s,t) \in S} E_{\Phi} \left(\frac{\partial^2 \ln f_{T_s T_t}(t_{is}, t_{it})}{\partial \phi_r \partial \phi_l} \right) \tag{9}$$

and $K(\Phi)$ by

$$K_{rl} = - \sum_{(s,t) \in S} E_{\Phi} \left(\frac{\partial \ln f_{T_s}(t_{is}, t_{it})}{\partial \phi_r} \frac{\partial \ln f_{T_l}(t_{is}, t_{it})}{\partial \phi_l} \right) \tag{10}$$

Similar in spirit to generalized estimating equations [24], this asymptotic normality result provides an easy way to estimate consistently the asymptotic covariance matrix. Indeed, the matrix J is found by evaluating the second derivate of the log $p\ell$ function at the PL estimate.

The expectation in K can be replaced by the cross-product of the observed scores. We will refer to J^{-1} as the model-based variance estimator, which should not be used as such because it overestimates precision; to K as the empirical correction; and $J^{-1}KJ^{-1}$ as the empirically corrected variance estimator.

A further advantage of the PL approach is the close connection of pseudo-likelihood with likelihood, enabling one to construct pseudo-likelihood ratio and pseudo-score test statistics that have easy-to-compute expressions and intuitively appealing distributions [25].

As discussed by Arnold and Strauss [23], the Cramèr–Rao inequality implies that $J^{-1}KJ^{-1}$ is greater than the inverse of I , corresponding to the Fisher information matrix for the maximum likelihood case, in the sense that $J^{-1}KJ^{-1} - I^{-1}$ is positive semidefinite. Therefore, a PL estimator is always less efficient than the corresponding ML estimator. Aerts *et al.* (2002) show that in many realistic settings efficiency losses are minor.

4. ASSOCIATION MEASURES

The Plackett–Dale model allows us to estimate and interpret the strength of the association between a pair of survival times via global cross-ratios (the θ parameters in the model). Therefore, θ may be considered a natural candidate for the measure of association. However, some researchers may feel that it is hard to get a feel for because it ranges throughout the entire real line. Further, different copulas (like the Clayton and Hougaard copulas) carry different and less straightforward association parameters. In such a situation it would be easier to work with a transformation of θ that has the interpretational properties of a correlation coefficient, such as Kendall's τ or Spearman's ρ . These will be discussed in turn.

4.1. Kendall's τ

Kendall's τ can be seen as the difference between the probability of concordance and the probability of discordance of two realizations of (T_1, T_2) . This coefficient lies in the $[-1, 1]$ interval and a zero value implies independence between T_1 and T_2 . There exists a relationship between Kendall's τ and θ for any copula $C(t_1, t_2, \theta)$ [26]:

$$\tau(\theta) = 4 \int_0^1 \int_0^1 C_{T_1 T_2}(t_1, t_2, \theta) C_{T_1 T_2}(dt_1, dt_2, \theta) - 1 \quad (11)$$

The marginal distributions of T_1 and T_2 do not affect (11), and hence it follows that τ only depends on the copula function $C_{T_1 T_2}$ [27]. Kendall's τ thus measures the association between both time points after adjustment for the covariates used in the model. Such a relationship is very simple for the Clayton and Hougaard copulas [11]. Precisely, one obtains $\tau = (\theta - 1)/(\theta + 1)$ for Clayton and $\tau = 1 - \theta$ for Hougaard. Estimates and confidence intervals (using the delta method) are accordingly easily obtained. There is no closed form for Kendall's τ in the Plackett–Dale case and an estimate has to be obtained directly from (11). We have developed a SAS IML 8.02 macro to this effect.

4.2. Spearman's ρ

Spearman's ρ is also based on concordance and discordance, independent of the marginal distributions, and belongs to the interval $[-1, 1]$. It can be shown that Spearman's ρ equals

Pearson's product-moment for grades of a pair of continuous random variables. The relationship between Spearman's ρ and the copula function is

$$\rho(\theta) = 12 \int_0^1 \int_0^1 C_{T_1, T_2}(t_1, t_2, \theta) dt_1 dt_2 - 3 \quad (12)$$

In contrast to the previous case, there is a closed-form expression in the Plackett–Dale case:

$$\rho(\theta) = \frac{\theta + 1}{\theta - 1} - \frac{2\theta \ln \theta}{(\theta - 1)^2} \quad (13)$$

An estimate follows from $\rho = \rho(\hat{\theta})$, with delta-method variance

$$\text{Var}(\hat{\rho}) = \left[\frac{-4(\hat{\theta} - 1) + 2(\hat{\theta} + 1) \ln \hat{\theta}}{(\hat{\theta} - 1)^3} \right]^2 \text{Var}(\hat{\theta})$$

From (13), the following asymptotic properties are derived: $\rho(\theta) \rightarrow 0$ when $\theta \rightarrow 1$, $\rho(\theta) \rightarrow -1$ when $\theta \rightarrow 0$ and $\rho(\theta) \rightarrow 1$ when $\theta \rightarrow \infty$.

5. CASE STUDIES

We are now in a position to analyse the data from Sections 2.1 and 2.2. Pseudo-likelihood estimates were obtained by using Newton–Raphson with analytical first derivatives and numerical second derivatives, implemented in SAS IML 8.02 and using routine NLPNRR (SAS Institute Inc. 1999–2001). Standard errors of the parameters were calculated using the inverse of the observed matrix of second derivatives. Although in these two examples a trivariate model is considered, the methodology is fully generally applicable to longer sequences of time-to-event endpoints. Indeed, the structure of the SAS programs allows us to fit any model and any number of outcomes with only minor changes. Using a flexible design matrix structure, a large class of model specifications is possible.

5.1. Analysis of the adoption study

We first consider bivariate analyses, selecting pairs out of the three possible survival times of interest. The first aim is to describe the biological associations between mother, father and child, and then to study the environmental effect, e.g., correlations with the adoptive parents. In each case, a trivariate analysis is envisaged. We will start with bivariate analyses and compare these results with those obtained from modelling the trivariate data directly. We will use the abbreviations BM, BF and ACh for biological mother, biological father in the biological models, replacing BM with AM and BF with AF in the adoptive models. The corresponding subscripts are 1, 2 and 3 in each case. All results for the biological families are presented in Table II, while Table III presents estimates for the adoptive families. The marginal distributions are all assumed to be Weibull with parameters λ_j and p_j , $j = 1, 2, 3$, and we consider three different parameters β_1 , β_2 , and β_3 to adjust for the sex of the child as it was done by Parner [10]. All association parameters are assumed to be constant.

It is clear from the way in which PL is defined that ML estimates are exactly the same when only two outcomes are considered. Although model-based standard errors and empirically

Table I. Adoption study: Model for the biological families.

Par.	BM-BF	BM-ACh	BF-ACh	BM-BF-ACh
θ_{12}	1.076(0.128;0.128)	—	—	1.076(0.127)
θ_{13}	—	1.164(0.193;0.187)	—	1.164(0.187)
θ_{23}	—	—	1.176(0.194;0.202)	1.175(0.201)
β_1	-0.085(0.086; 0.077)	-0.086(0.086; 0.077)	—	-0.084(0.069)
β_2	-0.009(0.078; 0.072)	—	-0.010(0.078; 0.072)	-0.004(0.036)
β_3	—	-1.066(0.164; 0.159)	-1.060(0.164; 0.159)	-1.063(0.137)
p_1	0.220(0.017;0.015)	0.219(0.017;0.015)	—	0.220(0.013)
p_2	0.279(0.011;0.010)	—	0.279(0.011;0.010)	0.279(0.006)
p_3	—	0.086(0.054;0.063)	0.085(0.054;0.063)	0.086(0.054)
λ_1	3.818(0.146;0.178)	3.817(0.146;0.179)	—	3.818(0.155)
λ_2	5.568(0.179;0.201)	—	5.568(0.179;0.200)	5.568(0.174)
λ_3	—	2.312(0.175;0.290)	2.313(0.176;0.291)	2.313(0.252)
τ_{12}	0.016(0.003,0.029)	—	—	0.016(0.003,0.029)
τ_{13}	—	0.034(0.016,0.051)	—	0.034(0.016,0.051)
τ_{23}	—	—	0.036(0.018,0.054)	0.036(0.017,0.054)
(Parner) τ_{12}	0.035(0.024,0.045)	—	—	—
(Parner) τ_{13}	—	0.050(0.036,0.064)	—	—
(Parner) τ_{23}	—	—	0.037(0.023,0.050)	—
ρ_{12}	0.024(-0.053, 0.102)	—	—	0.024(-0.053, 0.102)
ρ_{13}	—	0.051(-0.054, 0.155)	—	0.051(-0.054, 0.155)
ρ_{23}	—	—	0.054(-0.054, 0.162)	0.054(-0.058, 0.165)
(Parner) ρ_{12}	0.052(-0.010, 0.113)	—	—	—
(Parner) ρ_{13}	—	0.075(-0.010, 0.165)	—	—
(Parner) ρ_{23}	—	—	0.055(-0.027, 0.137)	—

Note: Maximum likelihood estimates (model-based standard errors; empirically corrected standard errors) of bivariate survival times and pseudo-likelihood estimates (standard errors) for trivariate model. For Kendall's τ and Spearman's ρ , estimates and 95% confidence intervals are given.

corrected standard errors i.e., those based on (8) are numerically different, they are of similar magnitudes and no clear ordering is seen between them. The tables reveal that the model-based standard errors calculated by means of the information matrix and the empirically corrected ones differ only slightly. Common parameters estimated using two different bivariate models are similar since all models are of a marginal type. By 'marginal type', we mean that the univariate marginal parameters in a bivariate model have exactly the same meaning as their counterparts in the corresponding univariate model. For example, $\hat{\beta}_1 = -0.085$ in model BM-BF as opposed to $\hat{\beta}_1 = -0.086$ in BM-ACh.

Tables I and II include all three types of association parameters: not only the log odds ratios θ but also Kendall's τ and Spearman's ρ , as introduced in Section 4. We observe that the association is not very strong but nevertheless significantly different from zero in some cases. The τ and ρ parameters are relatively similar but, in spite of them ranging on the same scale, they have a different meaning and they are not directly comparable.

Let us now turn attention to the trivariate situation. Let us consider a model with different association parameters for each pair of outcomes θ_{12} , θ_{13} and θ_{23} and with different parameters

Table II. Adoption study: Model for the biological and adoptive families.

Par.	AM-AF	AM-ACh	AF-ACh	AM-AF-ACh
θ_{12}	1.265(0.132;0.127)	—	—	1.265(0.127)
θ_{13}	—	0.844(0.138;0.133)	—	0.849(0.133)
θ_{23}	—	—	1.237(0.200;0.198)	1.240(0.197)
β_1	-0.015(0.077; 0.072)	-0.012(0.077; 0.072)	—	-0.029(0.064)
β_2	0.078(0.075;0.074)	—	0.077(0.075;0.074)	0.025(0.034)
β_3	—	-1.066(0.164; 0.159)	-1.064(0.164; 0.158)	-1.068(0.137)
p_1	0.210(0.009;0.009)	0.210 (0.009;0.009)	—	0.211(0.008)
p_2	0.235(0.008;0.008)	—	0.235(0.008;0.008)	0.241(0.005)
p_3	—	0.085(0.054;0.063)	0.085(0.054;0.063)	0.086(0.055)
λ_1	6.402(0.203;0.218)	6.406(0.203;0.219)	—	6.405(0.189)
λ_2	7.223(0.210;0.220)	—	7.228(0.210;0.022)	7.222(0.191)
λ_3	—	2.312(0.176;0.290)	2.311(0.176;0.291)	2.312(0.252)
τ_{12}	0.052(-0.045, 0.150)	—	—	0.052(0.041,0.063)
τ_{13}	—	-0.038(-0.184, 0.108)	—	-0.036(-0.060, -0.012)
τ_{23}	—	—	0.047(0.030,0.065)	0.048(0.030,0.065)
(Parner) τ_{12}	0.051(0.040,0.061)	—	—	—
(Parner) τ_{13}	—	-0.069(-0.085, -0.052)	—	—
(Parner) τ_{23}	—	—	0.041(0.027,0.054)	—
ρ_{12}	0.078(-0.501, 0.657)	—	—	0.078(0.013,0.143)
ρ_{13}	—	-0.057(-0.931, 0.818)	—	-0.055(-0.198, 0.089)
ρ_{23}	—	—	0.071(-0.033, 0.175)	0.072(-0.034, 0.177)
(Parner) ρ_{12}	0.076(0.013,0.140)	—	—	—
(Parner) ρ_{13}	—	-0.103(-0.202, -0.004)	—	—
(Parner) ρ_{23}	—	—	0.061(-0.021, 0.143)	—

Note: Maximum likelihood estimates (model-based standard errors; empirically corrected standard errors) of bivariate survival times and pseudo-likelihood estimates (standard errors) for trivariate model. For Kendall's τ and Spearman's ρ , estimates and 95% confidence intervals are given.

for the covariates corresponding to each outcome β_1 , β_2 and β_3 . Specific Weibull distributions with different scale and shape parameters for each outcome were used to model the marginals, i.e., p_1 , p_2 , p_3 , λ_1 , λ_2 and λ_3 . Effectively, this is the trivariate version of the previous bivariate ones. For the trivariate models, only empirically corrected standard errors are given in Tables I and II, since the model-based ones ignore the fact that in using all pairs out of three survival times on a cluster, all outcomes are used twice, leading to an exaggerated precision. Therefore, model-based standard errors are useless, even if all marginal and association models are correctly specified. We like to point out this feature since it is different from the GEE setting. Other than being a disadvantage, it is merely a 'side effect' of the way marginal pseudo-likelihood works. Let us add that obtaining convergence was not different and using different sets of starting values showed stability of the process.

Parameters retain the meaning they had in the bivariate models, with two advantages. First, using the data in a trivariate model is more efficient than using them in three separate models. Second, one avoids the occurrence of double estimates for the marginal parameters (β , λ , and

p parameters), in spite of them being not too different between various bivariate models. The same model was applied to the biological and adoptive families, enabling to contrast both sets of dependencies.

Comparisons of our association parameters with the ones given by Parner [10] cannot be made directly, since they are expressed on different scales. The association in our case is the global odds ratio, while Parner's quantity is based on the mean and variance of the assumed Gamma distribution. Therefore, both sets of association parameters are transformed to Kendall's τ and Spearman's ρ . There is a close agreement between both methods, and both enable consideration of multivariate models.

According to Parner's conclusions, the environmental association between the adoptive child and the mother was significant and negative; the environmental association between the adoptive father and the adoptive mother was significant. In our case, we can see from Table III that the estimated Kendall's coefficients are $\tau_{13} = -0.036$ with a 95% confidence interval $(-0.060, -0.012)$ and $\tau_{12} = 0.052$ with a 95% confidence interval $(0.041, 0.063)$, respectively. These results suggest that the longevity of the mother and the adoptive child were negatively correlated. Thus, we arrived at the same conclusions. The estimates are similar to the estimates obtained using Parner's model as shown in Table II. We could also test for equal environmental effects and genetic effects using a Wald-type test, but this is not the main goal of this work; details can be found in Reference [10].

5.2. Analysis of the AIDS study

In this section, we analyse the data described in Section 2.1. In the original paper by Finkelstein *et al.* [28] the pattern of the development of opportunistic infections in HIV-infected patients was evaluated, based on a cohort of 1530 patients. For the sake of illustration, we will work with a random sample of 1000 patients to reduce the computational burden. In principle it is not impossible to work with larger samples. The more common AIDS-defining opportunistic infections are *Pneumocystis carinii* pneumonia (PCP), *Mycobacterium avium* complex (MAC), cytomegalovirus (CMV) and systemic mycosis. These authors performed all the analysis adjusted for CD4 count. Without loss of generality, we perform the analysis for three time-to-event outcomes: PCP, CMV and the overall survival time of the AIDS patients (DTH). The main objective is to describe the association between all three outcomes after adjusting by CD4 count and treatment effect.

Parameters are subscripted with 1, 2 and 3 to refer to CMV, DTH and PCP, respectively. For the sake of illustration, consider β_T to be the *common* treatment effect and β_1 , β_2 and β_3 the outcome-specific parameters associated with the CD4 count. We will assume a Weibull distribution with parameters p_1 , p_2 , p_3 , λ_1 , λ_2 and λ_3 . Therefore, the vector of parameters to be estimated has 13 components:

$$\Phi = (\theta_{12}, \theta_{13}, \theta_{23}, \beta_T, \beta_1, \beta_2, \beta_3, p_1, p_2, p_3, \lambda_1, \lambda_2, \lambda_3) \quad (14)$$

where θ_{12} , θ_{13} and θ_{23} are the global cross-ratios. Using straightforward generalized linear models technology, it is straightforward to construct the overall design matrix \mathbf{X} , consisting of 13 columns (as many as there are parameters), and $3 \times 7 \times N$ rows. The calculation of the number of rows follows because there are 3 pairs to be formed out of three outcomes, for each pair (i.e. for each bivariate model), there are 7 "natural" parameters (an association parameter, and then a β , λ , and p parameter for each component of the pair). More details

Table III. AIDS study: Maximum likelihood estimates (model-based standard errors; empirically corrected standard errors) of bivariate survival times and pseudo-likelihood estimates (standard errors) for trivariate model.

Par.	CMV–DTH	CMV–PCP	DTH–PCP	CMV–DTH–PCP
θ_{12}	5.165(2.570;2.401)	—	—	4.369(1.165)
θ_{13}	—	4.434(1.850;2.182)	—	4.466(1.446)
θ_{23}	—	—	3.943(1.023;0.959)	3.691(0.865)
β_T	-0.054(0.020;0.020)	0.183(0.032;0.033)	-0.014(0.019;0.019)	0.016(0.111)
β_1	1.708(1.816;1.681)	1.504(1.892;1.547)	—	1.579(1.095)
β_2	2.160(0.706;0.752)	—	2.010(0.696;0.703)	2.069(0.732)
β_3	—	2.037(1.570;1.845)	2.168(1.487;1.838)	2.109(1.169)
p_1	-0.240(0.137;0.142)	-0.657(0.193;0.184)	—	-0.451(0.350)
p_2	0.341(0.033;0.038)	—	0.353(0.032;0.035)	0.338(0.164)
p_3	—	-1.147(0.257;0.331)	-0.807(0.203;0.270)	-0.958(0.469)
λ_1	1.606(0.033;0.030)	1.406(0.023;0.022)	—	1.487(0.136)
λ_2	1.941(0.015;0.017)	—	1.933(0.015;0.016)	1.940(0.111)
λ_3	—	1.117(0.012;0.014)	1.215(0.014;0.018)	1.161(0.108)
τ_{12}	0.352(0.307,0.397)	—	—	0.318(0.292,0.345)
τ_{13}	—	0.321(0.272,0.370)	—	0.323(0.291,0.355)
τ_{23}	—	—	0.297(0.273,0.322)	0.284(0.260,0.308)
ρ_{12}	0.503(0.269,0.736)	—	—	0.459(0.318,0.599)
ρ_{13}	—	0.462(0.204,0.721)	—	0.464(0.295,0.634)
ρ_{23}	—	—	0.430(0.298,0.563)	0.412(0.283,0.541)

Note: For Kendall's τ and Spearman's ρ , estimates and 95% confidence intervals are given.

on the design matrix are given in Appendix B. Generalization to more than three outcomes is straightforward and the SAS macro we developed carries the general situation. Parameter estimates are summarized in Table III.

Parameters in common between different bivariate models are generally fairly close, with the exception of β_T , which is even changing signs. While not significant, this is a clear indication that the trivariate model is the more appealing one, in spite of a larger standard error. Note that for some, but not all, parameters the standard error produced by the trivariate model is smaller. The log global cross-ratios θ are quite large, showing a strong association between all pairs of outcomes. Also here, Kendall's τ and Spearman's ρ are calculated to get a better grip on the association. Based on the correlation parameters ρ , a consistent picture of a correlation around 0.5 emerges.

6. CONCLUDING REMARKS

In this paper, we have extended the Plackett–Dale model for survival data to the multivariate case and we have shown that pseudo-likelihood estimation, in the sense of Arnold and Strauss [23], is a viable and attractive alternative to maximum likelihood in case of multivariate survival data. Maximum likelihood becomes prohibitive for large sequences of times, due to computational requirements. In contrast, the pseudo-likelihood procedure gives quite

satisfactory results. In addition, we proposed other association measures and we have shown the link of Spearman's ρ and Kendall's τ to the association parameter of the Plackett–Dale model θ . The method yields consistent and asymptotically normal estimates of the parameters of interest and the computational complexity is manageable.

The choice of the Plackett–Dale model was motivated by the fact that the association parameter θ has a natural interpretation for this copula. However, other copulas can be considered [12, 14, 29, 30]. To this end, checking the goodness of fit of copulas to bivariate survival data can be carried out by using the method proposed by Wang and Wells [31], and an adaptation of this method to our framework is a topic for future research. It is also worth noting that, while in this work we considered Weibull marginal distributions, it is possible to use other distributional assumptions, or even use a semi-parametric approach with unspecified baseline hazard functions [14].

The approach we presented gives a flexible tool for modelling any kind of time-to-event data accounting for the association between two or more outcomes. To illustrate our findings we have applied the proposed method in two different situations. Also, we have shown how the standard errors of the parameters need to be corrected in order to account for the lack of independence introduced by the fact that the information of a single subject is used more than once.

APPENDIX A

A1. Log likelihood function for the bivariate Plackett–Dale model

Let (T_1, T_2) denote paired failures times and (S_1, S_2) , (f_1, f_2) the corresponding marginal survival and density functions. Then, the joint survival and density functions of (T_1, T_2) are given by

$$S(t_1, t_2) = F_{T_1 T_2}(S_{T_1}(t_1), S_{T_2}(t_2)) \quad (\text{A1})$$

$$f(t_1, t_2) = \frac{\partial^2 S(t_1, t_2)}{\partial t_1 \partial t_2} f_{T_1}(t_1) f_{T_2}(t_2) \quad (\text{A2})$$

with $t_1, t_2 \geq 0$.

Let us denote by (C_1, C_2) the paired censoring times. For $i = 1, \dots, n$, assume that (T_{i1}, T_{i2}) and (C_{i1}, C_{i2}) are independent. For each i we observe $T_{ij} = \min(X_{ij}, C_{ij})$ $j = 1, 2$ then $\Delta_{ij} = I\{X_{ij} = T_{ij}\}$, i.e., indicates whether the lifetime is observed ($\Delta_{ij} = 1$) or not ($\Delta_{ij} = 0$).

We can write now the log likelihood function by combining the following different situations in one expression as follows:

Case	Δ_{i1}	Δ_{i2}
I	1	1
II	1	0
III	0	1
IV	0	0

Therefore,

$$\begin{aligned} \log \ell = & \sum_{i=1}^n \Delta_{i1} \Delta_{i2} \log(f(t_{i1}, t_{i2})) + \Delta_{i1}(1 - \Delta_{i2}) \log\left(-\frac{\partial S(t_{i1}, t_{i2})}{\partial t_1}\right) \\ & + (1 - \Delta_{i1}) \Delta_{i2} \log\left(-\frac{\partial S(t_{i1}, t_{i2})}{\partial t_2}\right) + (1 - \Delta_{i1})(1 - \Delta_{i2}) \log(S(t_{i1}, t_{i2})) \quad (\text{A3}) \end{aligned}$$

where $S(t_1, t_2)$ and $f(t_1, t_2)$ were defined in (A1) and (A2), respectively.

A2. Distribution function and its derivatives for $\theta \neq 1$

$$F(u, v, \theta) = \frac{1}{2(\theta - 1)} + \frac{u + v}{2} - \frac{H(u, v, \theta)}{2(\theta - 1)}$$

$$H(u, v, \theta) = \sqrt{(1 + (\theta - 1)(u + v))^2 - 4\theta(\theta - 1)uv}$$

$$\frac{\partial H}{\partial u} = \frac{(\theta - 1)}{H(u, v, \theta)} [(1 + (\theta - 1)(u + v)) - 2\theta v]$$

$$\frac{\partial H}{\partial v} = \frac{(\theta - 1)}{H(u, v, \theta)} [(1 + (\theta - 1)(u + v)) - 2\theta u]$$

$$\frac{\partial H}{\partial \theta} = \frac{(1 + (\theta - 1)(u + v))(u + v) - 2uv(2\theta - 1)}{H(u, v, \theta)}$$

$$\frac{\partial^2 H}{\partial u^2} = \frac{[(\theta - 1)^2 - (\partial H / \partial u)^2]}{H(u, v, \theta)}$$

$$\frac{\partial^2 H}{\partial v^2} = \frac{[(\theta - 1)^2 - (\partial H / \partial v)^2]}{H(u, v, \theta)}$$

$$\frac{\partial^2 H}{\partial \theta^2} = \frac{[(u - v)^2 - (\partial H / \partial \theta)^2]}{H(u, v, \theta)}$$

$$\frac{\partial^2 H}{\partial u \partial \theta} = \frac{\partial H}{\partial u} \left[\frac{1}{\theta - 1} - \frac{1}{H(u, v, \theta)} \frac{\partial H}{\partial \theta} \right] + \frac{(\theta - 1)(u - v)}{H(u, v, \theta)}$$

$$\frac{\partial^2 H}{\partial v \partial \theta} = \frac{\partial H}{\partial v} \left[\frac{1}{\theta - 1} - \frac{1}{H(u, v, \theta)} \frac{\partial H}{\partial \theta} \right] + \frac{(\theta - 1)(v - u)}{H(u, v, \theta)}$$

$$\frac{\partial^2 H}{\partial u \partial v} = -\frac{1}{H(u, v, \theta)} \left[\frac{\partial H}{\partial u} \frac{\partial H}{\partial v} + (\theta - 1)(\theta + 1) \right]$$

$$\begin{aligned} \frac{\partial F}{\partial u} &= \frac{1}{2} \left[1 - \frac{1}{\theta - 1} \frac{\partial H}{\partial u} \right] \\ \frac{\partial F}{\partial v} &= \frac{1}{2} \left[1 - \frac{1}{\theta - 1} \frac{\partial H}{\partial v} \right] \\ \frac{\partial F}{\partial \theta} &= -\frac{H(u, v, \theta)}{\theta - 1} + \frac{1}{2(\theta - 1)} \left[u + v - \frac{\partial H}{\partial \theta} \right] \\ \frac{\partial^2 F}{\partial u^2} &= -\frac{1}{2(\theta - 1)} \left(\frac{\partial H}{\partial u} \right)^2 \\ \frac{\partial^2 F}{\partial v^2} &= -\frac{1}{2(\theta - 1)} \left(\frac{\partial H}{\partial v} \right)^2 \\ \frac{\partial^2 F}{\partial \theta^2} &= -\frac{1}{\theta - 1} \left[2 \frac{\partial F}{\partial \theta} + \frac{1}{2} \frac{\partial^2 H}{\partial \theta^2} \right] \\ \frac{\partial^2 F}{\partial u \partial v} &= -\frac{1}{2(\theta - 1)} \left[\frac{\partial H}{\partial u} \frac{\partial H}{\partial v} + (\theta - 1)(\theta + 1) \right] \\ \frac{\partial^2 F}{\partial u \partial \theta} &= \frac{1}{2H(u, v, \theta)(\theta - 1)} \left[\frac{\partial H}{\partial u} \frac{\partial H}{\partial \theta} - (\theta - 1)(u - v) \right] \\ \frac{\partial^2 F}{\partial v \partial \theta} &= \frac{1}{2H(u, v, \theta)(\theta - 1)} \left[\frac{\partial H}{\partial v} \frac{\partial H}{\partial \theta} - (\theta - 1)(v - u) \right] \\ \frac{\partial^3 F}{\partial u^3} &= \frac{1}{H(u, v, \theta)} \frac{\partial H}{\partial u} \left[-\frac{\partial^2 F}{\partial u^2} + \frac{1}{\theta - 1} \frac{\partial^2 H}{\partial u^2} \right] \\ \frac{\partial^3 F}{\partial v^3} &= \frac{1}{H(u, v, \theta)} \frac{\partial H}{\partial v} \left[-\frac{\partial^2 F}{\partial v^2} + \frac{1}{\theta - 1} \frac{\partial^2 H}{\partial v^2} \right] \\ \frac{\partial^3 F}{\partial u^2 \partial v} &= \frac{1}{H(u, v, \theta)} \left[-\frac{\partial^2 F}{\partial u^2} \frac{\partial H}{\partial v} + \frac{1}{\theta - 1} \frac{\partial^2 H}{\partial v \partial u} \frac{\partial H}{\partial u} \right] \\ \frac{\partial^3 F}{\partial u \partial v^2} &= \frac{1}{H(u, v, \theta)} \left[\frac{\partial^2 F}{\partial v^2} \frac{\partial H}{\partial u} - \frac{1}{\theta - 1} \frac{\partial^2 H}{\partial u \partial v} \frac{\partial H}{\partial v} \right] \\ \frac{\partial^3 F}{\partial u^2 \partial \theta} &= \frac{1}{\theta - 1} \frac{\partial^2 F}{\partial u^2} + \frac{1}{H(u, v, \theta)} \left[-\frac{\partial^2 F}{\partial u^2} \frac{\partial H}{\partial \theta} + \frac{1}{\theta - 1} \frac{\partial H}{\partial u} \frac{\partial^2 H}{\partial u \partial \theta} \right] \\ \frac{\partial^3 F}{\partial u \partial \theta^2} &= -\frac{1}{\theta - 1} \frac{\partial^2 F}{\partial u \partial \theta} - \frac{1}{H(u, v, \theta)} - \frac{\partial^2 F}{\partial u \partial \theta} \frac{\partial H}{\partial \theta} \\ &\quad + \frac{1}{2H(u, v, \theta)(\theta - 1)} \left[\frac{\partial^2 H}{\partial u \partial \theta} \frac{\partial H}{\partial \theta} + \frac{\partial H}{\partial u} \frac{\partial^2 H}{\partial \theta^2} - (u - v) \right] \end{aligned}$$

$$\frac{\partial^3 F}{\partial u \partial v \partial \theta} = -\frac{1}{\theta - 1} \frac{\partial^2 F}{\partial u \partial v} - \frac{1}{H(u, v, \theta)} - \frac{\partial^2 F}{\partial u \partial v} \frac{\partial H}{\partial \theta} + \frac{1}{2H(u, v, \theta)(\theta - 1)} \left[\frac{\partial^2 H}{\partial u \partial \theta} \frac{\partial H}{\partial v} + \frac{\partial H}{\partial u} \frac{\partial^2 H}{\partial v \partial \theta} + 2\theta \right]$$

APPENDIX B

Let us exemplify the construction of a design matrix for the AIDS case study. The contribution of a single individual can be seen in our case as the contribution of three pseudo-likelihood individuals. Thus, \mathbf{X} can be written as N blocks,

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{pmatrix}$$

where the block corresponding to subject i is expressed as

$$\mathbf{X}_i = \begin{pmatrix} X_{i12} \\ X_{i13} \\ X_{i23} \end{pmatrix}$$

where

$$\mathbf{X}_{i12} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & trt_i & cd4_i & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & trt_i & 0 & cd4_i & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\mathbf{X}_{i13} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & trt_i & cd4_i & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & trt_i & 0 & 0 & cd4_i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\mathbf{X}_{i23} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & trt_i & 0 & cd4_i & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & trt_i & 0 & 0 & cd4_i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

ACKNOWLEDGEMENTS

The first and third authors wish to thank 'Bijzonder Onderzoeksfonds' of the Limburgs Universitair Centrum. We are grateful to Thorkild I.A. Sørensen, G.G Nielsen, P.K. Andersen, T.W Teasdale and C. Holst for their kind permission to use the adoption data. We thank the AIDS Clinical Trial Group (ACTG) for allowing to use the ACTG 116A and 116B/117 data. We acknowledge support from Interuniversity Attraction Poles Programme P5/24-Belgian State-Federal Office for Scientific, Technical and Cultural Affairs.

REFERENCES

1. Fleming TR, Harrington DP. *Counting Processes and Survival Analysis*. Wiley: New York, 1991.
2. Cox DR. Regression models in life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**:187–220.
3. Molenberghs G, Lesaffre E. Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association* 1994; **89**:633–644.
4. Dolin R, Amato D, Fischl MA, Pettinelli C, Beltangady M, Liou S, Brown MJ, Cross AP, Hirsch MS, Hardy WD, Mildvan D, Blair DC, Powderly WG, Para MF, Fife KH, Steigbigel RT, Smaldone L, the National Institute of Allergy and Infectious Diseases Clinical Trials Group. Zidovudine compared with didanosine in patients with advanced human immunodeficiency virus type I infection and little or no previous experience with zidovudine. *Archives of Internal Medicine* 1995; **155**:961–974.
5. Kahn JO, Lagakos SW, Richman DD, Cross A, Pettinelli C, Liou S, Brown M, Volberding PA, Crumacker CS, Beall G, Sacks HS, Merigan TC, Beltangady M, Smaldone L, Dolin R, the NIAID AIDS Clinical Trials Group. A controlled trial comparing continued zidovudine with didanosine in human immunodeficiency virus infection. *New England Journal of Medicine* 1992; **327**:581–587.
6. Saah AJ, Hoover DR, He Y, Kingsley LA, Phair JP, the Multicenter AIDS Cohort Study. Factors influencing survival after AIDS: Report from the Multicenter AIDS Cohort Study (MACS). *Journal of Acquired Immune Deficiency Syndromes* 1994; **7**:287–295.
7. Molenberghs G, Williams PL, Lipsitz SR. Prediction of survival and opportunistic infections in HIV infected patients: a comparison of imputation methods of incomplete CD4 counts. *Statistics in Medicine* 2002; **21**: 1387–1408.
8. Sørensen TIA, Nielsen GG, Andersen PK, Teasdale TW. Genetic and familial environmental influence on premature death of adults adoptees. *New England Journal of Medicine* 1988; **318**:727–732.
9. Nielsen GG, Gill RD, Andersen PK, Sørensen TIA. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* 1992; **19**:25–44.
10. Parner ET. A composite likelihood approach to multivariate survival data. *Scandinavian Journal of Statistics* 2001; **28**:295–302.
11. Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D. Validation of surrogate endpoints in multiple randomized clinical trials with failure-time endpoints. *Applied Statistics* 2001; **50**:405–422.
12. Oakes D. Bivariate survival models induced by frailties. *Journal of the American Statistical Association* 1989; **84**:487–493.
13. Lehmann EL. *Theory of Point Estimation*. Wiley: New York, 1983.

14. Shih JH, Louis TA. Inferences on association parameter in copula models for bivariate survival data. *Biometrics* 1995; **51**:1384–1399.
15. Plackett RL. A class of bivariate distributions. *Journal of the American Statistical Association* 1965; **60**: 516–522.
16. Fréchet M. Sur les tableaux de corrélation dont les marges sont données. *Annals Université de Lyon, Section A, Series 3* 1951; **14**:143–153.
17. Mardia KV. *Families of Bivariate Distributions*. Griffin: London, 1970.
18. Dale JR. Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* 1986; **42**:909–917.
19. Molenberghs G, Lesaffre E. Marginal modelling of multivariate categorical data. *Statistics in Medicine* 1999; **18**:2237–2255.
20. Le Cessie S, Van Houwelingen JC. Logistic regression for correlated binary data. *Applied Statistics* 1994; **43**: 95–108.
21. Geys H, Molenberghs G, Lipsitz SR. A note on the comparison of pseudo-likelihood and generalized estimating equations for marginal odds ratio models. *Journal of Statistical Computation and Simulation* 1998; **62**:45–72.
22. Geys H, Molenberghs G, Ryan L. Pseudo-likelihood modelling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association* 1999; **94**:34–745.
23. Arnold BC, Strauss D. Pseudo likelihood estimation: some examples. *Sankhya B* 1991; **53**:233–243.
24. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **50**:13–22.
25. Aerts M, Geys H, Molenberghs G, Ryan L. *Topics in Modelling of Clustered Data*. Chapman & Hall: London, 2002.
26. Genest C, McKay J. The joy of copulas: bivariate distributions with uniform marginals. *American Statistician* 1986; **40**:280–283.
27. Schweizer B, Wolff EF. On nonparametric measures of dependence for random variables. *Annals of Statistics* 1981; **9**:879–885.
28. Finkelstein DM, Williams PL, Nberghs G, Feinberg J, Powderly W, Kahn J, Dolins R, Cotton D. Patterns of opportunistic infections in patients with HIV infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 1996; **12**:38–45.
29. Joe H. *Multivariate Models and Dependence Concepts*. Chapman & Hall, London, 1997.
30. Nelsen RG. *An Introduction to Coulas*. Lecture Notes in Statistics, **139**. Springer: New York, 1999.
31. Wang W, Wells MT. Model selection and semiparametric inferences for bivariate failure-time data. *Journal of the American Statistical Association* 2000; **95**:62–76.