The Canadian Journal of Statistics
Vol. 42, No. 3, 2014, Pages 365–383
La revue canadienne de statistique

Semiparametric methods for survival analysis of case-control data subject to dependent censoring

Douglas E. SCHAUBEL[1]*, Hui ZHANG[2], John D. KALBFLEISCH[1] and Xu SHU[1]

[1]*Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA*
[2]*U.S. Food and Drug Administration, Silver Spring, MD, USA*

Key words and phrases: Case-control study; Cox regression; dependent censoring; estimating equation; inverse weighting

MSC 2010: Primary 62N01; secondary 62D05

Abstract: Case-control sampling can be an efficient and cost-saving study design, wherein subjects are selected into the study based on the outcome of interest. It was established long ago that proportional hazards regression can be applied to case-control data. However, each of the various estimation techniques available assumes that failure times are independently censored. Since independent censoring is often violated in observational studies, we propose methods for Cox regression analysis of survival data obtained through case-control sampling, but subject to dependent censoring. The proposed methods are based on weighted estimating equations, with separate inverse weights used to account for the case-control sampling and to correct for dependent censoring. The proposed estimators are shown to be consistent and asymptotically normal, and consistent estimators of the asymptotic covariance matrices are derived. Finite-sample properties of the proposed estimators are examined through simulation studies. The methods are illustrated through an analysis of pre-transplant mortality among end-stage liver disease patients obtained from a national organ failure registry. *The Canadian Journal of Statistics* 42: 365–383; 2014

Résumé: L'échantillonnage cas-témoins peut constituer un plan d'expérience efficace et économique dans le cadre duquel les sujets sont choisis pour l'étude en fonction du phénomène étudié. Il est établi depuis longtemps que le modèle de régression à risques proportionnels peut s'appliquer à des données cas-témoins. Cependant, toutes les techniques d'estimation existantes supposent que les temps de défaillance sont censurés de façon indépendante. Étant donné que l'indépendance de la censure est souvent bafouée dans le cadre d'études observationnelles, les auteurs proposent des méthodes pour la régression de Cox de données de survie sujettes à la censure dépendante obtenues par un échantillonnage cas-témoins. Les méthodes proposées se fondent sur des équations d'estimation pondérées dont les poids séparés et inverses permettent de tenir compte de l'échantillonnage cas-témoins et de corriger le biais lié à la censure dépendante. Les auteurs montrent que les estimateurs proposés sont convergents et asymptotiquement normaux. Ils obtiennent également des estimateurs convergents pour les matrices de covariance asymptotique. Ils examinent les propriétés de ces estimateurs sur des échantillons de taille finie par voie de simulation et illustrent les méthodes au moyen d'une analyse de données sur le taux de mortalité prétransplantation chez les patients atteints d'une maladie hépatique en phase terminale provenant d'un registre national d'organes défaillants. *La revue canadienne de statistique* 42: 365–383; 2014

1. INTRODUCTION

Case-control and case-cohort sampling schemes are designed to over-sample subjects expected to provide relatively large amounts of information in terms of parameter estimation. In the case-

* *Author to whom correspondence may be addressed.*
E-mail: deschau@umich.edu

control study, subjects observed to experience the event of interest (the cases) are over-sampled, while only a fraction of subjects not experiencing the event of interest (the controls) are selected. In the original formulation of case-control study, all cases are selected but, more generally, cases are over-sampled, such that their sampling fraction would be many times greater than that expected through random sampling. In the context of proportional hazards regression (Cox, 1972), the case-cohort design (Prentice, 1986) was proposed, in which one selects all cases, as well as a random sample of the study population (the sub-cohort). The case-cohort design is especially well-suited to Cox regression since the sub-cohort representing the risk set can be used to compute partial likelihood contributions at each observed event.

A number of methods have been proposed for the regression analysis of case-control and case-cohort studies under the proportional hazards model. Prentice & Breslow (1978) adapted the Cox model for use in case-control studies, and studied in detail the correspondence between Cox regression in the presence of prospective and retrospective sampling. Oakes (1981) provided a partial likelihood interpretation for the likelihood function used by Prentice & Breslow (1978) and, hence, a theoretical justification of the estimation procedure. Self & Prentice (1988) provided a rigorous theoretical evaluation of the case-cohort design of Prentice (1986), including detailed asymptotic derivations and various efficiency calculations. Kalbfleisch & Lawless (1988) developed pseudo-likelihood methods for semiparametric estimation of the illness-death model, applicable to case-cohort and other retrospective sampling designs. Lin & Ying (1993) considered general missing data problems in the context of Cox regression, with the case-cohort design cast as a special case. The asymptotic development by Lin & Ying (1993) leads to a different variance estimator for the Prentice (1986) case-cohort estimator than that derived by Self & Prentice (1988).

Much of the subsequent research on the case-cohort design was focussed on modifications to the sampling scheme and/or analysis in order to increase efficiency. For example, Chen & Lo (1999) derived a class of estimators for case-cohort sampling; essentially, the methods involve adding future cases to the risk set (in addition to the sub-cohort) in order to improve efficiency relative to the original method by Prentice (1986). Borgan et al. (2000) proposed stratified case-cohort designs to improve efficiency. Motivated by family-based studies, Moger, Pawitan, & Borgan (2008) developed case-cohort methods for fitting frailty models to large clustered data sets. The methods were inspired by the work of Kalbfleisch & Lawless (1988) and Borgan et al. (2000). The motivation for the work of Moger, Pawitan, & Borgan (2008) was to reduce computational burden, rather than dollar cost.

With respect to the case-control study, much of the development in the last 15 years has been directed at clustered data. For instance, in the case-control family study, one selects a random sample of cases, random sample of controls, as well as the family members of the cases and controls. Li, Yang, & Schwartz (1998) developed parametric methods for fitting copula models to clustered survival data obtained through case-control family sampling. Shih & Chatterjee (2002) extended this work to the semiparametric setting through quasi-likelihood procedures. Hsu & Gorfine (2006) developed pseudo-likelihood methods for fitting frailty model to case-control family data using an approach similar to that of Shih & Chatterjee (2002). Hsu et al. (2004) developed semiparametric methods for frailty models based on case-control family sampling, with the focus being on the conditional regression parameter. Gorfine, Zucker, & Hsu (2009) further evaluated frailty models based on case-control family data through a pseudo-full likelihood approach, and derived rigorous large-sample theory.

There have been several evaluations of cohort sampling in a general sense. Chen & Lo (1999) described the close theoretical connection between the case-cohort and case-control study within the framework of Cox regression. Chen (2001) considered "generalized case-cohort sampling" (nested case-control, case-cohort and case-control, all within the same framework) in the context of Cox regression. Gray (2009) studied various cohort sampling designs, focusing on Inverse

Probability of Selection Weighted (IPSW) estimators applied to the Cox model (e.g., Binder, 1992; Barlow, 1994; Borgan et al., 2000; Lin, 2000; Pan & Schaubel, 2008; Nan, Kalbfleisch, & Yu, 2009).

The motivation for subsampling of large epidemiologic databases is well-established in the literature. Nothing prevents such data from being subject to dependent censoring. However, each of the subsampling methods cited thus far (and, to the best of our knowledge, all existing methods for Cox regression analysis of case-control data) assumes that subjects are censored in a manner independent of the failure rate. In most practical applications in which the independent censoring assumption is violated, censoring is actually a mixture of independent (e.g., administrative, or random loss to follow-up) and dependent (e.g., censoring of pre-treatment death by the initiation of treatment). In such cases, it is necessary to distinguish between subjects who were independently and dependently censored. The most popular method for handling dependent censoring is Inverse Probability of Censoring Weighting (IPCW) proposed by Robins & Rotnitzky (1992). From this perspective, case-control studies appear to extend nicely to the dependent censoring setting. In particular, since IPCW entails modelling the dependent censoring hazard, there is an incentive to over-sample dependently censored subjects.

We propose semiparametric methods for the analysis of failure time data generated by case-control sampling and subject to dependent censoring. In the setting of interest, we over-sample dependently censored subjects, since correcting for dependent censoring through IPCW requires that the dependent censoring process be modelled. The dependent censoring hazard is modelled using existing case-control survival analysis methods featuring Inverse Probability of Sampling Weighting (IPSW). Parameter estimation for the failure time hazard model then proceeds through weighted estimating equations, with the weights correcting for both the case-control sampling and the dependent censoring. We derive asymptotic properties of the proposed estimator, in a manner which accounts for the randomness in the estimated IPCW and IPSW components. Since the derived asymptotic variance is complicated, we suggest an alternative variance estimator that is much easier to program and much faster to compute. Finite-sample performance of the proposed procedures is evaluated through simulation.

In the evaluation of a proposed subsampling method, it is useful to assess not only if the method can be implemented in a real data set, but, in addition, whether the method gives reasonable answers in practice. Along these lines, we apply the proposed methods to an analysis of pre-transplant survival among end-stage liver disease (ESLD) patients. In this application, the full cohort is available, meaning that the results based on the full-cohort can serve as a target for the results obtained through the proposed case-control sampling methods. With respect to background, liver transplantation is the preferred method of treatment. However, there are thousands more patients needing a liver transplant than there are available donor organs. As such, patients clinically suitable for receiving a deceased-donor liver transplant are put on a wait-list, with the pertinent time origin ($t = 0$) then being the date of wait listing. The receipt of a liver transplant does not censor death but, naturally, does censor pre-transplant death (i.e., the time until death in the absence of liver transplantation). In liver transplantation, patients are ranked on the wait-list based on their most recent Model for End-stage Liver Disease (MELD) score. MELD was chosen as the scoring system largely because it is a very strong predictor of pre-transplant mortality (e.g., Wiesner et al., 2001; Kremers et al., 2004; Huo et al., 2005; Merion et al., 2005; Basto et al., 2008; Subramanian et al., 2010). Since MELD is time-dependent, an analysis of the effect of baseline risk factors (i.e., measured at time $t = 0$) on the pre-transplant death hazard could result in substantial bias if liver transplantation were treated as independent censoring. This and other related issues in the liver transplant setting are discussed by Schaubel et al. (2009).

In certain cases, special exceptions are made under which a wait-listed patient may be assigned a MELD score which is higher than that calculated, in an attempt to reflect the patient's perceived

medical urgency. The most frequent occurrence of such MELD exceptions is for patients with hepatocellular carcinoma (HCC), a form of liver cancer. HCC patients are typically assigned a MELD score of 22, which is often considerably higher than the score based on their laboratory measures. To our knowledge, no existing analyses in the liver transplant literature have quantified whether the MELD score of 22 accurately reflects the true "MELD-equivalent" wait-list mortality risk faced by HCC patients. As a primary example in this article, we compare HCC versus non-HCC patients, with the latter group categorized by baseline MELD score. Since MELD (at time 0 but, also, after time 0) affects both death and liver transplantation probabilities, liver transplantation is handled as dependent censoring of pre-transplant death time.

It should be noted that MELD exception scores for HCC are usually assigned at initial wait listing (time 0). Therefore, given our analytic objective and in the interests of interpretation, it is appropriate to compare HCC patients to non-HCC patients, with the latter categorized specifically by time 0 MELD score. To use MELD at time > 0 would be tantamount to having the HCC and comparator patients established at different times, which would render the resulting parameter estimates uninterpretable. More generally, the perils of adjusting for time-dependent factors (such as MELD at time >0) are well-described in the causal inference literature (e.g., Hernán, Brumback, & Robins, 2000). Another general setting for which adjustment for time-dependent factors is explicitly avoided concerns the setting wherein there is a mediator. In evaluating a baseline factor (e.g., treatment assignment), adjusting for the mediator (e.g., though a time-dependent indicator covariate) would generally distort the resulting treatment effect; it might then be necessary to treat the mediator as dependent censoring.

The remainder of this article is organized as follows. In Section 2, we describe the proposed estimation procedures. In Section 3, we derive large sample properties for the proposed estimators. We conduct simulation studies in Section 4 to investigate the finite-sample properties of the proposed estimators. Section 5 provides an application of the methods to the afore-described end-stage liver disease data. The article concludes with a discussion in Section 6.

2. METHODS

Let $\mathbf{Z}_{1i}$ denote the q_1-vector of time-constant covariates for subject i ($i = 1, \ldots, n$). Let $\mathbf{Z}_{2i}(t)$ be the q_2-vector of time-dependent covariates at time t, $\mathbf{Z}_i(t) = \{\mathbf{Z}_{1i}^T, \mathbf{Z}_{2i}(t)^T\}^T$, and $\widetilde{\mathbf{Z}}_i(t) = \{\mathbf{Z}_i(u) : 0 \le u \le t\}$ denote the history of $\mathbf{Z}_i(\cdot)$ up to time t. Let T_i and C_i be the potential failure and censoring times, respectively. We suppose that $C_i = C_{1i} \wedge C_{2i}$, where $a \wedge b = \min\{a, b\}$, C_{1i} is the censoring time due to mechanisms that are independent of T_i given $\mathbf{Z}_i(0)$, and C_{2i} denotes the dependent censoring time; that is, C_{2i} is dependent on T_i given $\mathbf{Z}_i(0)$. Let $X_i = T_i \wedge C_i$, $Y_i(t) = I(X_i \ge t)$, $\Delta_{1i} = I(T_i \le C_i)$, $\Delta_{2i} = I(C_{2i} \le C_{1i}, C_{2i} < T_i)$, $\Delta_{3i} = (1 - \Delta_{1i})(1 - \Delta_{2i})$, $N_i(t) = I(X_i \le t, \Delta_{1i} = 1)$ and $N_i^C(t) = I(X_i \le t, \Delta_{2i} = 1)$, where $I(\cdot)$ is the indicator function. The observable data are assumed to be n independently and identically distributed copies of $\{N_i(\cdot), N_i^C(\cdot), Y_i(\cdot), \mathbf{Z}_i(\cdot)\}$. Let ξ_i indicate whether or not subject i is sampled. The variate ξ_i is allowed to depend on Δ_{1i}, Δ_{2i} and Δ_{3i} so that the sampling probability can be different for subjects who fail, subjects who are dependently censored and those who are independently censored. Let the cohort be divided into three strata according to the outcome $(\Delta_1, \Delta_2, \Delta_3)$ such that $L_k = \{i : \Delta_{ki} = 1\}$, $k = 1, 2, 3$. Let $p_k = \text{pr}(\xi_i = 1 \mid i \in L_k)$, $\boldsymbol{p} = (p_1, p_2, p_3)^T$ and $\rho_i(\boldsymbol{p}) = \sum_{k=1}^3 \Delta_{ki}\xi_i / p_k$. Note that $\rho_i(\boldsymbol{p})$ weights the ith subject by the inverse probability that the subject is sampled.

We assume that the hazard of failure for individual i is specified by the following proportional hazards model (Cox, 1972),

$$\lambda_i\{t \mid \mathbf{Z}_i(0)\} = \lambda_0(t)\exp\{\boldsymbol{\beta}_0^T\mathbf{Z}_i(0)\}, \tag{1}$$

 DOI: 10.1002/cjs

where $\lambda_0(t)$ is an unspecified baseline hazard function for failure time, and $\boldsymbol{\beta}_0$ is a $(q_1 + q_2)$-dimensional regression parameter. Note that we are chiefly interested in inferring the role of $\mathbf{Z}_i(0)$ on the failure time hazard, as opposed to $\{\mathbf{Z}_i(t) : t > 0\}$, for reasons of interpretation. For example, it is straightforward to predict survival probability from time 0 using a pre-specified value of $\mathbf{Z}_i(0)$ along with parameter estimates from model (1). To do so using a model based on $\widetilde{\mathbf{Z}}_i(t)$ would be much more complicated, and would generally require modelling the stochastic properties of the process $\mathbf{Z}_i(t)$.

If it were also the case that C_{2i} was independent of T_i given $\mathbf{Z}_i(0)$ (unlike the setting of interest), then $\boldsymbol{\beta}_0$ could be consistently estimated by $\widehat{\boldsymbol{\beta}}$, the root of the estimating equation $\mathbf{U}(\boldsymbol{\beta}) = 0$, where

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \rho_i(\boldsymbol{p})\{\mathbf{Z}_i(0) - \overline{\mathbf{Z}}(\boldsymbol{\beta}, t)\} dN_i(t), \tag{2}$$

where $\tau < \infty$ is the maximum follow-up time, $\overline{\mathbf{Z}}(\boldsymbol{\beta}, t) = \mathbf{S}^{(1)}(\boldsymbol{\beta}, t)/S^{(0)}(\boldsymbol{\beta}, t)$, $\mathbf{S}^{(d)}(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} \rho_i(\boldsymbol{p}) Y_i(t) \mathbf{Z}_i(0)^{\otimes d} \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(0)\}$, with $\mathbf{a}^{\otimes 0} = 1, \mathbf{a}^{\otimes 1} = \mathbf{a}$, and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$. Estimating equations of the same general structure as (2) and arising from IPSW have been proposed by several previous authors, for example, Kalbfleisch & Lawless (1988), Binder (1992), Borgan et al. (2000) and Lin (2000) for the Cox model; Kulich & Lin (2000) for the additive hazards model; and Nan, Kalbfleisch, & Yu (2009) for the accelerated failure time model.

Since $\mathbf{Z}_i(t)$ affects both the event and censoring times, and since $\mathbf{Z}_i(t)$ is not incorporated into model (1), C_{2i} would generally not be independent of T_i given $\mathbf{Z}_i(0)$. In this case, the estimate $\widehat{\boldsymbol{\beta}}$ derived from (2) could be substantially biased because (2) does not accommodate the dependence between C_{2i} and T_i. We assume that conditional on the covariate history $\widetilde{\mathbf{Z}}_i(t)$, the hazard of dependent censoring C_{2i} at time t does not further depend on the possibly unobserved failure time T_i, that is,

$$\lambda_i^C\{t \mid \widetilde{\mathbf{Z}}_i(T_i), C_i \geq t, T_i \geq t, T_i\} = \lambda_i^C\{t \mid \widetilde{\mathbf{Z}}_i(t), C_i \geq t, T_i \geq t\}. \tag{3}$$

This fundamental assumption is called "no unmeasured confounders for censoring" (Rubin, 1977; Robins, 1993). Borrowing terminology from the competing risks literature (Kalbfleisch and Prentice, 2002), assumption (3) allows us to identify the cause-specific hazard for C_{2i}. We assume a time-dependent Cox proportional hazards model for the right-hand side of Equation (3),

$$\lambda_i^C\{t \mid \widetilde{\mathbf{Z}}_i(t), X_i \geq t\} = \lambda_0^C(t) \exp\{\boldsymbol{\alpha}_0^T \mathbf{V}_i(t)\}, \tag{4}$$

where $\lambda_0^C(t)$ is an unspecified baseline hazard function for dependent censoring, $\mathbf{V}_i(t)$ is a s-vector consisting of functions of $\widetilde{\mathbf{Z}}_i(t)$, and $\boldsymbol{\alpha}_0$ is a s-dimensional regression parameter.

We propose the following estimating function,

$$\mathbf{U}_R(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} R_i(t)\{\mathbf{Z}_i(0) - \overline{\mathbf{Z}}_R(\boldsymbol{\beta}, R, t)\} dN_i(t), \tag{5}$$

where

$$\overline{\mathbf{Z}}_R(\boldsymbol{\beta}, R, t) = \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, R, t)}{S^{(0)}(\boldsymbol{\beta}, R, t)}$$

$$\mathbf{S}^{(d)}(\boldsymbol{\beta}, R, t) = n^{-1} \sum_{i=1}^{n} R_i(t) Y_i(t) \mathbf{Z}_i(0)^{\otimes d} \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(0)\}$$

$$R_i(t) = \rho_i(\boldsymbol{p}) W_i(t)$$
$$W_i(t) = e^{\Lambda_i^C(t)} \kappa_i(t),$$

where $\Lambda_i^C(t) = \int_0^t \exp\{\boldsymbol{\alpha}^T \mathbf{V}_i(u)\} d\Lambda_0^C(u)$ and the function $\kappa_i(t)$ in the weight $W_i(t)$ is a stabilization factor. We consider three choices of $\kappa_i(t)$. One choice is $\kappa_{1i}(t) = 1$. However, when the censoring is heavy, $e^{\Lambda_i^C(t)}$ could be quite large and lead to instability in the estimation. In this case, the choice of $\kappa_{2i}(t) = \exp[-\int_0^t \exp\{\boldsymbol{\alpha}^T \mathbf{V}_i(0)\} d\Lambda_0^C(u)]$ or $\kappa_{3i}(t) = \exp[-\Lambda_i^\dagger\{t \mid \mathbf{Z}_i(0)\}]$ may be more appropriate, where $\Lambda_i^\dagger(t)$ is based on a time-to-censoring model that uses only the baseline covariate values, $\mathbf{Z}_i(0)$. Hereafter, we denote $W_{ji}(t) = e^{\Lambda_i^C(t)} \kappa_{ji}(t)$, $j = 1, 2, 3$, and correspondingly estimate $\boldsymbol{\beta}_0$ with $\widehat{\boldsymbol{\beta}}_{W_1}$, $\widehat{\boldsymbol{\beta}}_{W_2}$ and $\widehat{\boldsymbol{\beta}}_{W_3}$, the solutions to $\mathbf{U}_R(\boldsymbol{\beta}) = 0$ with weights $W_{1i}(t)$, $W_{2i}(t)$ and $W_{3i}(t)$, respectively.

The weight $W_{1i}(t)$ can be estimated using $\exp\{\widehat{\Lambda}_i^C(t)\}$, where

$$\widehat{\Lambda}_i^C(t) = \int_0^t \exp\{\widehat{\boldsymbol{\alpha}}^T \mathbf{V}_i(s)\} d\widehat{\Lambda}_0^C(s, \widehat{\boldsymbol{\alpha}})$$
$$\widehat{\Lambda}_0^C(t, \boldsymbol{\alpha}) = \sum_{i=1}^n \int_0^t \left[\sum_{j=1}^n \rho_j(\boldsymbol{p}) Y_j(s) \exp\{\boldsymbol{\alpha}^T \mathbf{V}_j(s)\} \right]^{-1} \rho_i(\boldsymbol{p}) dN_i^C(s),$$

where $\widehat{\boldsymbol{\alpha}}$, the partial likelihood estimate of $\boldsymbol{\alpha}_0$, is the root of $\mathbf{U}^C(\boldsymbol{\alpha}) = 0$, where

$$\mathbf{U}^C(\boldsymbol{\alpha}) = \sum_{i=1}^n \int_0^\tau \{\mathbf{V}_i(t) - \overline{\mathbf{V}}(\boldsymbol{\alpha}, \boldsymbol{p}, t)\} \rho_i(\boldsymbol{p}) dN_i^C(t),$$

is an IPSW-based estimating function, with $\overline{\mathbf{V}}(\boldsymbol{\alpha}, \boldsymbol{p}, t) = \mathbf{S}_C^{(1)}(\boldsymbol{\alpha}, \boldsymbol{p}, t) / S_C^{(0)}$ $(\boldsymbol{\alpha}, \boldsymbol{p}, t)$ and $\mathbf{S}_C^{(d)}(\boldsymbol{\alpha}, \boldsymbol{p}, t) = n^{-1} \sum_{i=1}^n \rho_i(\boldsymbol{p}) Y_i(t) \mathbf{V}_i(t)^{\otimes d} e^{\boldsymbol{\alpha}^T \mathbf{V}_i(t)}$.

The weight $W_{2i}(t)$ can be estimated using $\widehat{\kappa}_{2i}(t) \exp\{\widehat{\Lambda}_i^C(t)\}$, where $\widehat{\kappa}_{2i}(t) = \exp[-\widehat{\Lambda}_i^C\{t, \widehat{\boldsymbol{\alpha}} \mid \mathbf{Z}_i(0)\}]$; that is, $\kappa_{2i}(t)$ is estimated using the same model (4), but only using baseline covariate values,

$$\widehat{\Lambda}_i^C\{t, \widehat{\boldsymbol{\alpha}} \mid \mathbf{Z}_i(0)\} = \int_0^t \exp\{\widehat{\boldsymbol{\alpha}}^T \mathbf{V}_i(0)\} d\widehat{\Lambda}_0^C(s, \widehat{\boldsymbol{\alpha}}).$$

The weight $W_{3i}(t)$ can be estimated by $\widehat{\kappa}_{3i}(t) \exp\{\widehat{\Lambda}_i^C(t)\}$, where $\widehat{\kappa}_{3i}(t) = \exp\{-\widehat{\Lambda}_i^\dagger(t)\}$, with $\kappa_{3i}(t)$ estimated using an additional baseline model for C_{2i},

$$\lambda_i^\dagger\{t \mid \mathbf{Z}_i(0), C_i \geq t, T_i, T_i \geq t\} = \lambda_0^\dagger(t) \exp\{\boldsymbol{\alpha}^{\dagger T} \mathbf{V}_i(0)\},$$

such that we have

$$\widehat{\Lambda}_i^\dagger(t) = \int_0^t \exp\{\widehat{\boldsymbol{\alpha}}^{\dagger T} \mathbf{V}_i(0)\} d\widehat{\Lambda}_0^\dagger(s, \widehat{\boldsymbol{\alpha}}^\dagger),$$
$$\widehat{\Lambda}_0^\dagger(t, \boldsymbol{\alpha}^\dagger) = \sum_{i=1}^n \int_0^t \left[\sum_{j=1}^n \rho_j(\boldsymbol{p}) Y_j(s) \exp\{\boldsymbol{\alpha}^{\dagger T} \mathbf{V}_j(0)\} \right]^{-1} \rho_i(\boldsymbol{p}) dN_i^C(s),$$

and $\widehat{\boldsymbol{\alpha}}^{\dagger}$ is the partial likelihood estimate of $\boldsymbol{\alpha}^{\dagger}$ under the model for dependent censoring with hazard $\lambda_i^{\dagger}(t)$. Weight stabilizers analogous to $\kappa_{3i}(t)$ have been suggested, for example, by Robins & Finkelstein (2000) and Hernán, Brumback, & Robins (2000). We propose the stabilizer $\kappa_{2i}(t)$ as an alternative. The performance of each of $W_{1i}(t)$, $W_{2i}(t)$ and $W_{3i}(t)$ are compared through simulations studies described in Section 4.

To summarize, $\widehat{\boldsymbol{\beta}}_{W_j}$ $(j = 1, 2, 3)$ is computed as the root of the estimating equation in (5), with $R_i(t)$ replaced by $\widehat{R}_i(t)$. After estimating $\boldsymbol{\beta}_0$, the cumulative baseline hazard, $\Lambda_0(t)$, can then be estimated by

$$\widehat{\Lambda}_0^W(t) = \sum_{i=1}^{n} \int_0^t \left[\sum_{\ell=1}^{n} \widehat{R}_\ell(s) Y_\ell(s) \exp\{\widehat{\boldsymbol{\beta}}_{W_j}^T \mathbf{Z}_\ell(0)\} \right]^{-1} \widehat{R}_i(s) dN_i(s).$$

3. ASYMPTOTIC PROPERTIES

The following conditions are assumed throughout this section.

(*a*) $\{N_i(\cdot), N_i^C(\cdot), Y_i(\cdot), \mathbf{Z}_i(\cdot)\}$, $i = 1, \ldots, n$ are independently and identically distributed.

(*b*) $P\{Y_i(\tau) = 1\} > 0$.

(*c*) $|Z_{ij}(0)| + \int_0^\tau |dZ_{ij}(t)| < B_Z < \infty$, where Z_{ij} is the jth component of $\mathbf{Z}_i$ and B_Z is a constant.

(*d*) There exists a neighbourhood $\mathcal{B}$ of β_0 such that $\sup_{u\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \|\mathbf{S}^{(d)}(\boldsymbol{\beta}, R, u) - \mathbf{s}^{(d)}(\boldsymbol{\beta}, R, u)\| \longrightarrow 0$ in probability for $d = 0, 1, 2$, where $\mathbf{s}^{(d)}(\boldsymbol{\beta}, R, u) = E\left\{\mathbf{S}^{(d)}(\boldsymbol{\beta}, R, u)\right\}$ is absolutely continuous, for $\boldsymbol{\beta} \in \mathcal{B}$, uniformly in $u \in (0, \tau]$, $E(\cdot)$ denotes expectation. Moreover, $s^{(0)}(\boldsymbol{\beta}, R, u)$ is assumed to be bounded away from zero.

(*e*) There exists a neighbourhood $\mathcal{B}_\mathcal{C}$ of $\boldsymbol{\alpha}_0$ such that $\sup_{u\in[0,\tau],\boldsymbol{\alpha}\in\mathcal{B}_\mathcal{C}} \|\mathbf{S}_C^{(d)}(\boldsymbol{\alpha}, \mathbf{p}, u) - \mathbf{s}_C^{(d)}(\boldsymbol{\alpha}, u)\| \longrightarrow 0$ in probability for $d = 0, 1, 2$, where $\mathbf{s}_C^{(d)}(\boldsymbol{\alpha}, u) = E\{\mathbf{S}_C^{(d)}(\boldsymbol{\alpha}, \mathbf{p}, u)\}$ is absolutely continuous, for $\boldsymbol{\alpha} \in \mathcal{B}_\mathcal{C}$, uniformly in $u \in (0, \tau]$. Moreover, $s_C^{(0)}(\boldsymbol{\alpha}, u)$ is assumed to be bounded away from zero.

(*f*) The matrices $\mathbf{A}(\boldsymbol{\beta}_0)$ and $\mathbf{A}^C(\boldsymbol{\alpha}_0)$ are positive definite, where

$$\mathbf{A}(\boldsymbol{\beta}) = \int_0^\tau \left\{ \mathbf{s}^{(2)}(\boldsymbol{\beta}, R, u)/s^{(0)}(\boldsymbol{\beta}, R, u) - \overline{\mathbf{z}}(\boldsymbol{\beta}, R, u)^{\otimes 2} \right\} dF(u)$$

$$\mathbf{A}^C(\boldsymbol{\alpha}) = \int_0^\tau \left\{ \mathbf{s}_C^{(2)}(\boldsymbol{\alpha}, u)/s_C^{(0)}(\boldsymbol{\alpha}, u) - \overline{\mathbf{v}}(\boldsymbol{\alpha}, u)^{\otimes 2} \right\} dF^C(u)$$

with $\overline{\mathbf{z}}(\boldsymbol{\beta}, R, u) = \mathbf{s}^{(1)}(\boldsymbol{\beta}, R, u)/s^{(0)}(\boldsymbol{\beta}, R, u)$, $\overline{\mathbf{v}}(\boldsymbol{\alpha}, u) = \mathbf{s}_C^{(1)}(\boldsymbol{\alpha}, u)/s_C^{(0)}(\boldsymbol{\alpha}, u)$, $F(u) = E\{R_i(u)N_i(u)\}$, $F^C(u) = E\left\{\rho_i(p_0)N_i^C(u)\right\}$.

(*g*) $\Lambda_0(\tau) < \infty$, $\Lambda_0^C(\tau) < \infty$.

We describe the asymptotic properties of the proposed estimators in the following theorems.

Theorem 1. *Under conditions (a) − (g), as $n \to \infty$, $n^{1/2}\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right)$ converges to a mean zero Normal distribution with covariance $\mathbf{A}^C(\boldsymbol{\alpha}_0)^{-1}\boldsymbol{\Omega}(\boldsymbol{\alpha}_0)\mathbf{A}^C(\boldsymbol{\alpha}_0)^{-1}$, where $\mathbf{A}^C(\boldsymbol{\alpha}_0)$ is as defined by Condition (f) and $\boldsymbol{\Omega}(\boldsymbol{\alpha}) = E\{\boldsymbol{\psi}_i(\boldsymbol{\alpha}, \mathbf{p})^{\otimes 2}\}$, with $\boldsymbol{\psi}_i(\boldsymbol{\alpha}, \mathbf{p})$ being asymptotically independent and identically distributed for $i = 1, \ldots, n$; we defer the definition of $\boldsymbol{\psi}_i(\boldsymbol{\alpha}, \mathbf{p})$ to the Supplementary Materials document.*

In the Supplementary Materials document, we show that $n^{1/2}\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right) = \mathbf{A}^C(\boldsymbol{\alpha}_0)^{-1} n^{-1/2} \sum_{i=1}^n \boldsymbol{\psi}_i(\boldsymbol{\alpha}_0, \mathbf{p}_0) + o_p(1)$; hence, $n^{1/2}\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right)$ is essentially a scaled sum of n independent and identically distributed random quantities with mean zero and finite variance. By the Multivariate Central Limit Theorem (MCLT) and empirical process theory, the asymptotic normality is proved.

Theorem 2. *Under conditions* $(a) - (g)$*, as* $n \to \infty$, $n^{1/2}\left(\widehat{\boldsymbol{\beta}}_{W_1} - \boldsymbol{\beta}_0\right)$*, converges to a mean zero Normal distribution with covariance* $\mathbf{A}(\boldsymbol{\beta}_0)^{-1}\boldsymbol{\Sigma}(\boldsymbol{\beta}_0, R)\mathbf{A}(\boldsymbol{\beta}_0)^{-1}$*, with* $\mathbf{A}(\boldsymbol{\beta})$ *having been defined in Condition (f) and where* $\boldsymbol{\Sigma}(\boldsymbol{\beta}, R) = E\{\boldsymbol{\Theta}_i(\boldsymbol{\beta}, R)^{\otimes 2}\}$*, with* $\boldsymbol{\Theta}_i(\boldsymbol{\beta}, R)$ *being independent and identically distributed mean 0 variates* $(i = 1, \ldots, n)$ *asymptotically. The explicit definition of* $\boldsymbol{\Theta}_i(\boldsymbol{\beta}, R)$ *is provided in the Supplementary Materials.*

The proof of Theorem 2 (provided in the Supplementary Materials) begins by decomposing $n^{1/2}\{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\}$ into $n^{1/2}\{\widehat{\Lambda}_0^C(t; \widehat{\boldsymbol{\alpha}}, \widehat{\mathbf{p}}) - \widehat{\Lambda}_0^C(t; \widehat{\boldsymbol{\alpha}}, \mathbf{p}_0)\} + n^{1/2}\{\widehat{\Lambda}_0^C(t; \widehat{\boldsymbol{\alpha}}, \mathbf{p}_0) - \widehat{\Lambda}_0^C(t; \boldsymbol{\alpha}_0, \mathbf{p}_0)\} + n^{1/2}\{\widehat{\Lambda}_0^C(t; \boldsymbol{\alpha}_0, \mathbf{p}_0) - \Lambda_0^C(t)\}$. Then $n^{1/2}\{\widehat{\Lambda}_0^C(t) - \Lambda_0^C(t)\}$ can be expressed asymptotically as a sum of independent and identically distributed zero-mean variates, as $n \to \infty$. Combining this result and the Functional Delta Method, we can show that $n^{1/2}\{\widehat{R}_i(t) - R_i(t)\}$ can be written asymptotically as a sum of independent and identically distributed zero-mean variates, as $n \to \infty$. Finally, through the Functional Delta Method, the asymptotic normality of $n^{1/2}(\widehat{\boldsymbol{\beta}}_{W_1} - \boldsymbol{\beta}_0)$ is obtained.

The expression for the asymptotic covariance of $\widehat{\boldsymbol{\beta}}_{W_1}$ is very complicated and difficult to implement numerically. A practical way to estimate the variance of the proposed estimators is to treat the weights $R_i(t)$ as known rather than estimated. In the setting where the weight function is known, results derived in the Supplementary Materials show that

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathbf{A}(\boldsymbol{\beta}_0)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \mathbf{U}_i^{\ddagger}\left\{\boldsymbol{\beta}_0, R\right\} + o_p(1) \tag{6}$$

with $\mathbf{U}_i^{\ddagger}(\boldsymbol{\beta}_0, R) = \int_0^\tau \{\mathbf{Z}_i(0) - \overline{\mathbf{z}}(\boldsymbol{\beta}, R, t)\} R_i(t) dM_i(t)$ and $dM_i(t) = dN_i(t) - Y_i(t) d\Lambda_i(t)$. Hence, $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically a scaled sum of independent and identically distributed zero-mean random quantities with finite variance. Therefore, the variance of $\widehat{\boldsymbol{\beta}}_{W_1}$ is estimated by $\widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}})^{-1}\widehat{\boldsymbol{\Sigma}}^{\ddagger}(\widehat{\boldsymbol{\beta}}, \widehat{R})\widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}})^{-1}$, where $\boldsymbol{\Sigma}^{\ddagger}(\boldsymbol{\beta}, R) = E\{\mathbf{U}_i^{\ddagger}(\boldsymbol{\beta}, R)^{\otimes 2}\}$, $\widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}})$ and $\widehat{\boldsymbol{\Sigma}}^{\ddagger}(\widehat{\boldsymbol{\beta}}, \widehat{R})$ are calculated by replacing limiting values with their corresponding empirical counterparts.

By similar arguments, the asymptotic normality holds for $n^{1/2}(\widehat{\boldsymbol{\beta}}_{W_2} - \boldsymbol{\beta}_0)$ and $n^{1/2}(\widehat{\boldsymbol{\beta}}_{W_3} - \boldsymbol{\beta}_0)$. However, the covariance will be even more complicated than that of $n^{1/2}(\widehat{\boldsymbol{\beta}}_{W_1} - \boldsymbol{\beta}_0)$. Therefore, similarly, we can treat $R_i(t)$ as fixed to calculate the variance of $\widehat{\boldsymbol{\beta}}_{W_2}$ and $\widehat{\boldsymbol{\beta}}_{W_3}$. Note that (6) holds when using W_2 or W_3, such that the variances of each of $\widehat{\boldsymbol{\beta}}_{W_2}$ and $\widehat{\boldsymbol{\beta}}_{W_3}$ can be estimated by $\widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}})^{-1}\widehat{\boldsymbol{\Sigma}}^{\ddagger}(\widehat{\boldsymbol{\beta}}, \widehat{R})\widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}})^{-1}$, with $R_i(t)$ being replaced by $\rho_i(\widehat{\mathbf{p}})\widehat{W}_{2i}(t)$ and $\rho_i(\widehat{\mathbf{p}})\widehat{W}_{3i}(t)$, respectively.

Our final asymptotic result pertains to the proposed cumulative baseline hazard function estimator.

Theorem 3. *Under conditions* $(a) - (g)$*,* $n^{1/2}\{\widehat{\Lambda}_0^W(t) - \Lambda_0(t)\}$ *converges to a zero-mean Gaussian process as* $n \to \infty$*, with an explicit covariance function estimator.*

A proof of Theorem 3 is provided in the Supplementary Materials, including definitions pertinent to the limiting covariance function.

TABLE 1: Simulation study: data configurations.

Parameter	1A	1B	2A	2B	3A	3B
λ_0	0.1	0.1	0.1	0.1	0.05	0.05
β_1	0.406	0.406	0.406	0.406	−0.406	−0.406
β_2	0.406	0.406	0.406	0.406	0.406	0.406
λ_0^C	0.1	0.1	0.05	0.05	0.05	0.05
α_1	−0.5	−0.5	−0.406	−0.406	−0.406	−0.406
α_2	0.693	0.693	1.010	1.010	1.010	1.010
α_3	0.095	0.095	0.095	0.095	0.095	0.095
$\%T$	51%	51%	62%	62%	27%	27%
$\%C_2$	44%	44%	28%	28%	45%	45%
$\%C_1$	5%	5%	10%	10%	28%	28%
corr(T, C_2)	0.13	0.13	0.20	0.20	0.27	0.27
Full-cohort	2,500	2,500	2,500	2,500	2,500	2,500
p_1	0.15	0.10	0.15	0.10	0.15	0.10
p_2	0.15	0.10	0.15	0.10	0.15	0.10
p_3	0.15	0.10	0.15	0.10	0.15	0.10
Case-control	375	250	375	250	375	250

4. SIMULATION STUDY

We investigated the finite-sample properties of the proposed estimators through a series of simulation studies. In each setting, the study population (i.e., full-cohort; prior to case-control sampling), was $n = 2{,}500$. The baseline covariate, to be used in the failure time hazard model, consisted of two independent Bernoulli variates, $\mathbf{Z}_i(0) = [Z_{1i}, Z_{2i}(0)]^T$, where Z_{1i} does not change over time. Each of Z_{1i} and $Z_{2i}(0)$ took the value 1 with probability 0.5. The independent censoring times C_{1i} were constant and equal to 12. After generating U_T from a Uniform(0,1) distribution, the failure time T was generated from a Cox model with hazard function

$$\lambda_i(t) = \lambda_0 \ \exp\{\beta_1 Z_{1i} + \beta_2 Z_{2i}(0)\}, \tag{7}$$

by solving the equation $\int_0^T \lambda_0(u) \exp\{\beta_1 Z_{1i} + \beta_2 Z_{2i}(0)\}\, du = -\log U_T$ for T, so that U_T corresponds to the survival function at T. The time-dependent covariate $Z_{2i}(t)$ was generated as

$$\begin{aligned} &Z_{2i}(0) I(U_T \le 0.3) + \{Z_{2i}(0) + U_T \times \text{int}(t)\} I(0.3 < U_T \le 0.6) \\ &\quad + \{Z_{2i}(0) + U_T/2 \times \text{int}(t)\} I(U_T > 0.6), \end{aligned} \tag{8}$$

where int(t) is the integer part of t. The dependent censoring time C_{2i} was then generated from a Cox model with hazard function

$$\lambda_i^C(t) = \lambda_0^C \ \exp[\alpha_1 Z_{1i} + \alpha_2 Z_{2i}(t) Z_{2i}(0) + \alpha_3 Z_{2i}(t)\{1 - Z_{2i}(0)\}]. \tag{9}$$

We evaluated three data (full-cohort) configurations and, within each, two different case-control sampling schemes. For each of the six scenarios, 1,000 replicates were generated. Parameter specifications are given in Table 1. Note that the $W_i(t)$ component of the weight was bounded by 10 in all runs.

TABLE 2: Simulation results based on 1,000 replications: setting 1A–1B.

$W_i(t)$	Estimator	BIAS	ESD	ASE	CP
		Setting 1A: $p_1 = p_2 = p_3 = 0.15$			
1	$\widehat{\beta}_1$	0.059	0.152	0.152	0.928
W_1	$\widehat{\beta}_1^{W_1}$	0.054	0.169	0.169	0.930
W_2	$\widehat{\beta}_1^{W_2}$	0.015	0.152	0.154	0.955
W_3	$\widehat{\beta}_1^{W_3}$	0.014	0.150	0.152	0.957
1	$\widehat{\beta}_2$	−0.087	0.150	0.149	0.905
W_1	$\widehat{\beta}_2^{W_1}$	−0.040	0.171	0.168	0.930
W_2	$\widehat{\beta}_2^{W_2}$	−0.003	0.151	0.153	0.948
W_3	$\widehat{\beta}_2^{W_3}$	−0.002	0.150	0.152	0.950
		Setting 1B: $p_1 = p_2 = p_3 = 0.10$			
1	$\widehat{\beta}_1$	0.061	0.190	0.187	0.933
W_1	$\widehat{\beta}_1^{W_1}$	0.060	0.213	0.206	0.925
W_2	$\widehat{\beta}_1^{W_2}$	0.021	0.190	0.189	0.951
W_3	$\widehat{\beta}_1^{W_3}$	0.020	0.188	0.187	0.947
1	$\widehat{\beta}_2$	−0.085	0.187	0.183	0.922
W_1	$\widehat{\beta}_2^{W_1}$	−0.041	0.212	0.203	0.927
W_2	$\widehat{\beta}_2^{W_2}$	−0.005	0.185	0.187	0.947
W_3	$\widehat{\beta}_2^{W_3}$	−0.004	0.184	0.186	0.948

By the specifications above, the time-dependent covariate $Z_{2i}(t)$ is correlated with the event time T_i through Equation (8). In addition, $Z_{2i}(t)$ also affects the censoring time C_{2i} via model (9). Since only the baseline value of $Z_{2i}(t)$, that is, $Z_{2i}(0)$, is adjusted for in the model (7), the censoring time C_{2i} is dependent on T_i given Z_{1i} and $Z_{2i}(0)$.

For each data generation, we selected failures, dependent censoring subjects and independent censoring subjects by Bernoulli sampling, with $p_1 = p_2 = p_3 = 0.15$ (Settings 1A, 2A, 3A) which, on average, resulted in a case-control sample $n = 375$ subjects; or $p_1 = p_2 = p_3 = 0.10$ (Settings 1B, 2B, 3B) which resulted in an average of $n = 250$ subjects.

For each scenario, we report bias; empirical standard deviation (ESD) across the 1,000 replicates; average asymptotic standard error (ASE); and empirical coverage probability (CP) for 95% confidence intervals. Four estimators were evaluated, which differ with respect to the handling of the IPCW component. The first estimator did not make an adjustment for dependent censoring ($W_i(t) = 1$). The second, third and fourth estimators are the three proposed estimators, with W_1 being based on the unstabilized correction for dependent censoring; while W_2 and W_3 refer to the different versions of the proposed stabilized estimators based on $W_{2i}(t)$ or $W_{3i}(t)$, respectively.

Results of the simulation study are provided in Tables 2, 3 and 4. Some general conclusions are as follows. The estimator which did not account for dependent censoring is notably biased, in all configurations. In some but not all cases, the bias is partly corrected by the estimator using $W_{1i}(t)$ (i.e., the unstabilized estimator). In certain cases, the bias is actually more pronounced than for the $W_i(t) = 1$ estimator. The bias due to dependent censoring is largely corrected by the stabilized versions of the proposed estimator, with the ASE being similar to the ESD and, correspondingly,

TABLE 3: Simulation results based on 1,000 replications: setting 2A–2B.

$W_i(t)$	Estimator	BIAS	ESD	ASE	CP
		Setting 2A: $p_1 = p_2 = p_3 = 0.15$			
1	$\widehat{\beta}_1$	0.096	0.136	0.139	0.901
W_1	$\widehat{\beta}_1^{W_1}$	0.090	0.149	0.151	0.911
W_2	$\widehat{\beta}_1^{W_2}$	0.042	0.140	0.146	0.953
W_3	$\widehat{\beta}_1^{W_3}$	0.028	0.136	0.143	0.954
1	$\widehat{\beta}_2$	−0.104	0.133	0.134	0.876
W_1	$\widehat{\beta}_2^{W_1}$	−0.082	0.150	0.147	0.911
W_2	$\widehat{\beta}_2^{W_2}$	−0.034	0.138	0.141	0.937
W_3	$\widehat{\beta}_2^{W_3}$	−0.021	0.135	0.139	0.947
		Setting 2B: $p_1 = p_2 = p_3 = 0.10$			
1	$\widehat{\beta}_1$	0.101	0.169	0.171	0.916
W_1	$\widehat{\beta}_1^{W_1}$	0.100	0.182	0.184	0.916
W_2	$\widehat{\beta}_1^{W_2}$	0.053	0.172	0.177	0.947
W_3	$\widehat{\beta}_1^{W_3}$	0.040	0.168	0.174	0.955
1	$\widehat{\beta}_2$	−0.102	0.167	0.164	0.912
W_1	$\widehat{\beta}_2^{W_1}$	−0.083	0.184	0.179	0.926
W_2	$\widehat{\beta}_2^{W_2}$	−0.036	0.171	0.171	0.944
W_3	$\widehat{\beta}_2^{W_3}$	−0.023	0.168	0.170	0.948

coverage probabilities approximating the nominal level. Of the three proposed estimators, the stabilized version based on $W_{3i}(t)$ has the best performance in the simulation studies.

5. APPLICATION

We applied the proposed methods to analyse pre-transplant mortality for patients with end-stage liver disease (ESLD). Data were obtained from the Scientific Registry of Transplant Recipients (SRTR). The study population consisted of patients who were initially wait-listed for liver transplantation in the United States at age ≥ 18 between March 1, 2002 and December 31, 2008. Patients were followed from the date of initial wait-listing until the earliest of death, receiving liver transplantation, loss to follow-up, or last day of the observation period (December 31, 2008). The time scale was in days. The primary outcome of interest is pre-transplant mortality. Loss to follow-up, living-donor transplantation and administrative censoring were considered to be independent censoring. Dependent censoring occurred through deceased-donor liver transplantation.

The Model of End-stage Liver Disease (MELD) score is time-dependent and is updated based on a frequency that ranges from weekly to yearly and depends on the last reported MELD. In the current liver allocation system, patients are ordered on the wait-list primarily by descending MELD. That is, patients with higher MELD are considered to be at greater medical urgency and, therefore, get higher priority for transplantation. However, for hepatocellular carcinoma (HCC) patients, the calculated MELD based on laboratory measures has generally been thought to understate actual medical urgency. As such, a MELD score of 22 is usually assigned to an

TABLE 4: Simulation results based on 1,000 replications: setting 3A–3B.

$W_i(t)$	Estimator	BIAS	ESD	ASE	CP
		$p_1 = p_2 = p_3 = 0.15$			
1	$\widehat{\beta}_1$	0.065	0.200	0.202	0.945
W_1	$\widehat{\beta}_1^{W_1}$	0.119	0.235	0.231	0.909
W_2	$\widehat{\beta}_1^{W_2}$	0.044	0.220	0.215	0.938
W_3	$\widehat{\beta}_1^{W_3}$	0.015	0.207	0.205	0.952
1	$\widehat{\beta}_2$	−0.260	0.216	0.206	0.757
W_1	$\widehat{\beta}_2^{W_1}$	−0.173	0.248	0.240	0.881
W_2	$\widehat{\beta}_2^{W_2}$	−0.065	0.241	0.230	0.935
W_3	$\widehat{\beta}_2^{W_3}$	−0.029	0.231	0.222	0.943
		$p_1 = p_2 = p_3 = 0.10$			
1	$\widehat{\beta}_1$	0.064	0.252	0.248	0.935
W_1	$\widehat{\beta}_1^{W_1}$	0.120	0.296	0.280	0.906
W_2	$\widehat{\beta}_1^{W_2}$	0.045	0.275	0.262	0.926
W_3	$\widehat{\beta}_1^{W_3}$	0.017	0.259	0.252	0.944
1	$\widehat{\beta}_2$	−0.259	0.264	0.255	0.825
W_1	$\widehat{\beta}_2^{W_1}$	−0.170	0.297	0.292	0.905
W_2	$\widehat{\beta}_2^{W_2}$	−0.064	0.288	0.280	0.941
W_3	$\widehat{\beta}_2^{W_3}$	−0.028	0.279	0.273	0.955

HCC patient if the laboratory MELD is less than 22. The primary objective of our analysis is to determine which range of (calculated) MELD score is actually consistent with the HCC pre-transplant mortality hazard.

The factor of interest is, in a general sense, underlying liver disease at wait listing (the time origin), classified as hepatocellular carcinoma (HCC) versus not. The objective is to determine the MELD score category to which pre-transplant HCC mortality is equivalent. Therefore, as will be detailed later in this section, the pre-transplant death model will have HCC as the reference category, to which all non-HCC patients (classified by baseline MELD category) are compared. The use of baseline (i.e., $t = 0$) MELD score is consistent with HCC diagnosis being applied at baseline. Note that the use of baseline versus time-dependent MELD depends on the analytic objectives. In Section 6, we describe a setting where use of time-dependent MELD is indicated.

In many studies, it has been shown that MELD is the dominant risk factor for liver pre-transplant mortality. Moreover, as stated in the previous paragraph, MELD also strongly affects the liver transplant hazard. Therefore, since the death model does not adjust for time-dependent MELD, pre-transplant mortality and deceased-donor liver transplantation will be correlated, such that it is necessary to account for liver transplantation as dependent censoring.

It is necessary to account for geography in our analysis. There are 60 Organ Procurement Organizations (OPO) in the United States, each of which maintains a liver transplant wait-list pertaining to a population size of a state, on average. Adjustment for OPO is important since it reflects geography, which can affect mortality in ways not reflected by the remainder of the adjustment covariates. With respect to transplantation, the transplant hazard differs markedly by

TABLE 5: Sampling design for the analysis of liver pre-transplant mortality.

Patients	OPO size	Deaths	Transplants	Censored	Total
		Full-cohort			
HCC	All	418	3,511	546	4,475
non-HCC	≤400	418	1,499	490	2,407
	401–1,300	3,676	10,816	5,230	19,722
	>1,300	6,076	12,794	10,477	29,347
Total		10,588	28,620	16,743	55,951
		Sampling fractions			
HCC	All	1	1	1	
non-HCC	≤400	1	1	1	
	401–1,300	0.30	0.10	0.10	
	>1,300	0.15	0.10	0.10	
		Case-control Sample			
HCC	All	418	3,511	546	4,475
non-HCC	≤400	418	1,499	490	2,407
	401–1,300	1,079	1,075	496	2,650
	>1,300	893	1,295	1,018	3,206
Total		2,808	7,380	2,550	12,738

OPO owing in part to differences in organ availability. For both the time-to-death and time-to-transplant models, we adjusted for OPO through stratification, since it may not be appropriate to assume proportionality with respect to the 60 OPO-specific baseline transplant or death hazard functions.

Specifics regarding the full cohort ($n = 55{,}951$), sampling fractions and case-control sample are provided in Table 5.

Since a relatively small fraction of the full cohort was diagnosed with HCC (8%; 4,475 patients), HCC patients were over-sampled. In addition, we wanted to have sufficient representation from OPOs of various sizes in the case-control sample. The way we classified the size of the OPO was based on the number of patients wait-listed in that OPO in the full-cohort. Random sampling within end-point would lead to inadequate representation of patients from smaller OPOs. Therefore, we assigned greater (lesser) sampling fractions to the small (large) OPOs. After Bernoulli sampling, the case-control sample consisted of 12,738 patients.

There are additional issues regarding the data structure which must be taken into account. In particular, a patient who is too sick to receive a transplant can be inactivated (usually a temporary measure) or removed (permanent) from the wait-list. During these intervals, the patient is ineligible to receive a transplant. Therefore, an appropriate Cox model in this setting is given by the following,

$$\lambda_{ij}^{C}(t) = A_i(t)\lambda_{0j}^{C}(t)\exp\{\boldsymbol{\alpha}^T\mathbf{V}_i(t)\}, \tag{10}$$

TABLE 6: Analysis comparing HCC (assigned MELD of 22; reference) versus non-HCC (by lab MELD) pre-transplant mortality.

	Case-control Unweighted: $W = 1$				Case-control Weighted: W_2			Full-cohort Weighted: W_2		
Patients	MELD	$\hat{\beta}$	SE	p	$\hat{\beta}$	SE	p	$\hat{\beta}$	SE	p
HCC	"22"	0	–	–	0	–	–	0	–	–
non-HCC	6–8	−1.07	0.12	<0.0001	−1.02	0.15	<0.0001	−0.81	0.09	<0.0001
	9–11	−0.60	0.09	<0.0001	−0.48	0.11	<0.0001	−0.50	0.08	<0.0001
	12–13	−0.53	0.09	<0.0001	−0.40	0.11	0.0002	−0.24	0.08	0.002
	14–15	−0.25	0.09	0.005	**0.002**	**0.11**	**0.98**	**0.01**	**0.08**	**0.88**
	16–17	**0.10**	**0.10**	**0.32**	0.24	0.12	0.0498	0.24	0.08	0.004
	18–19	0.26	0.12	0.02	0.63	0.15	<0.0001	0.52	0.09	<0.0001
	20–22	0.55	0.11	<0.0001	0.83	0.12	<0.0001	0.81	0.09	<0.0001
	23–24	0.90	0.17	<0.0001	1.41	0.21	<0.0001	1.12	0.11	<0.0001
	25–29	1.59	0.16	<0.0001	1.93	0.19	<0.0001	1.44	0.11	<0.0001
	30–39	2.38	0.16	<0.0001	2.69	0.17	<0.0001	1.86	0.12	<0.0001
	40	3.68	0.29	<0.0001	3.65	0.37	<0.0001	1.79	0.20	<0.0001

where j denotes OPO number ($j = 1, \ldots, 60$) and $A_i(t)$ is an indicator of being active on the wait-list (i.e., not being deactivated or previously removed) as of time t. The time-dependent covariate vector $\mathbf{V}_i(t)$ includes MELD at time t (grouped into intervals: [6,8], [9,11], [12,13], [14,15], [16,17], [18,19], [20,22], [23,24], [25,29], [30,39] and 40) with HCC patients chosen as the reference group. These MELD categories have been suggested by several previous analyses; the use of a categorical version of MELD negates the need to pin down its precise functional form (which is known to not be linear). The vector $\mathbf{V}_i(t)$ also includes the following baseline covariates: age group, gender, race and blood type, with age less than 40, Female, Caucasian and blood type O as reference levels, respectively. Note that, for the intervals where the patient was either inactivated or removed, the transplant hazard was treated as 0, as indicated by equation (10). As such, we delete patient subintervals with $A_i(t) = 0$ to fit model (10). However, since the inactivated or removed patients are still at risk of pre-transplant death, patient subintervals with $A_i(t) = 0$ are re-included in the fitting of the pre-transplant mortality model.

The pre-transplant death model is given by

$$\lambda_{ij}(t) = \lambda_{0j}(t) \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(0)\}, \tag{11}$$

where j again represents OPO and $\mathbf{Z}_i(0)$ contains indicators for the afore-listed MELD categories (HCC serving as the reference), and the adjustment covariates used in the transplant model (10). Since the IPCW weights could be very large toward the tail of the observation time, we truncated IPCW weights with 10.

Results of the analysis are shown in Table 6. We present results for two case-control analyses: the "unweighted" analysis ($W_i(t) = 1$), in which dependent censoring (liver transplantation) is treated as independent censoring; and the proposed weighted analysis, which corrects for dependent censoring through $W_{2i}(t)$. We also carried out a full-cohort analysis, again based on $W_{2i}(t)$. Table 6 shows that when the dependent censoring is treated as independent, HCC appears to

have pre-transplant mortality equal to that of MELD group 16–17 (see bolded entries in Table 6). However, accounting for the dependent censoring, HCC is the pre-transplant mortality equivalent of the MELD 14–15 group, a result consistent with the full-cohort analysis.

Based on the simulation studies in Section 4, the unstabilized weight $W_{1i}(t)$ was out-performed by both $W_{2i}(t)$ and $W_{3i}(t)$ in terms of empirical bias and variance. In our application, results based on $W_{1i}(t)$, $W_{2i}(t)$ and $W_{3i}(t)$ were all quite similar. It appeared from the simulations that $W_{3i}(t)$ performed somewhat better than $W_{2i}(t)$; in each case a reduction in bias (for $W_{3i}(t)$ relative to $W_{2i}(t)$) was accompanied by a concomitant reduction in standard deviation. However, in the full-cohort analysis, standard errors for $W_{2i}(t)$ tended to be less than those for $W_{3i}(t)$. Hence, we chose $W_{2i}(t)$ for the main line analysis. As mentioned previously, the IPCW component of the weight is susceptible to unduly large values, which implies the use of a cap. We employed a cap of 10, consistent with the simulation study. In the case-cohort analyses, the cap affected 16,507 records for $W_{1i}(t)$ (1.55% of the total 1,066,630 records used); 2,616 records (0.25%) for $W_{2i}(t)$ and 3,351 (0.31%) for $W_{3i}(t)$. Full cohort and case-control results based on $W_{1i}(t)$ and $W_{3i}(t)$ are provided in the Supplementary Materials.

The bias in the analysis which does not account for dependent censoring ($W_i(t) = 1$) is substantial, particularly upon consideration of the objectives of our analysis. This is clear upon comparing the estimators based on $W_i(t) = 1$ and $W_{2i}(t)$, either through relative bias, or simply the absolute difference in point estimates relative to the standard error. By either of these measures, the bias is considerable, except for the two most extreme categories: MELD 6–8 and MELD=40. Note that these two extreme categories are of least interest, since it was widely anticipated that the lowest (highest) MELD category would have much lower (higher) pre-transplant mortality risk than HCC. For the unweighted analysis, the HCC-mortality-equivalent is given by MELD 16–17, whereas for the weighted analysis, it appears that MELD 14–15 and HCC have equal mortality. From this perspective, the difference between the unweighted and weighted analysis would be considered very important to the liver transplant field.

6. DISCUSSION

In this article, we propose methods for proportional hazards modelling of survival data subject to dependent censoring and obtained through case-control sampling. The proposed methods employ a double-inverse-weighting scheme, through which the proposed estimators adjust for the sampling bias and overcome dependent censoring. Simulation studies show that the proposed estimators are approximately unbiased and that our asymptotic results are applicable to finite samples. The proposed estimates can be computed using standard software that accommodates left-truncation, weighting and offsets (e.g., PROC PHREG in SAS; coxph in R).

The case-control sampling scheme we consider involves either sampling all observed failures (cases), or a randomly selected fraction. The censored subjects (controls) are categorized as either censored dependently, or independently. There is an incentive to over-sample dependently censored subjects, such that one can estimate the IPCW weights with sufficient precision. Although the sampling scheme we study involves random sampling based only on final status (dead, censored), the methods extend in a straightforward fashion to the setting in which sampling also depends on fixed covariates, a property illustrated through the application to the liver data on Section 5. Although we consider case-control sampling explicitly, the proposed methods also apply if instead case-cohort sampling is utilized.

One could also view the sampling scheme considered in this article to be a form of outcome-dependent sampling (ODS). In an ODS design, one collects covariate information from a sample by allowing selection probabilities to depend on individuals' outcomes (e.g., death, survival). An ODS design concentrates resources on observations carrying the greatest amount of information.

There is a large literature on analysing data arising from ODS; see, for example, Breslow & Holubkov (1997a), Zhou et al. (2002), Schildcrout & Heagerty (2008), Song, Zhou, & Kosorok (2009) and Wang et al. (2009). It would be possible to generalize the proposed methods to the ODS setting, interpreting the version we develop in this article as a censored-data-specific instance of inverse-weighted generalized estimating equations.

We propose three different weights to correct the bias induced by dependent censoring. In general, when the dependent censoring is light or moderate, the unstabilized weight $W_1(t)$ works well. However, when censoring is heavy, $W_1(t)$ may be quite large toward the tail of the observation time resulting in unstable estimates. In this case, stabilized weights, $W_2(t)$ and $W_3(t)$, may be preferable and usually result in a less biased and more precise estimator than that arising from using the unstabilized weight $W_1(t)$. With respect to the stabilized weights, we found that $W_3(t)$ tended to out-perform $W_2(t)$, although not uniformly. An apparent advantage of $W_2(t)$ is that only one dependent censoring hazard model is required. However, details regarding the programming and ability to exploit functions currently available (in SAS or R) lead to $W_3(t)$ actually being the easier of the two stabilized weights to compute, even though a second hazard regression model is required. Overall, it would appear that $W_3(t)$ would generally be the preferred weight, although the stabilized weight that actually performs best is likely to be application-dependent. For instance, in our analysis of the liver disease data, $W_2(t)$ resulted in parameter estimators with much smaller standard errors than $W_3(t)$, leading to our preference for the former in this particular case.

In simulation studies, we treated the IPCW weights and IPSW weights as fixed to simplify the computation, which would result in conservative covariance estimators because those weights are actually estimated as opposed to being known. However, simulation results suggest that the proposed ASEs by treating the IPCW weights and IPSW weights as fixed are quite accurate. The simplification of the variance calculation would likely lead to invalid inference when the variance of the estimated weight was substantial, which would be in cases where the subsample was not large enough or did not contain a sufficient number of failures or dependently censored subjects. In such cases, the variability due to estimating the weight parameters would not be a negligible fraction of the total variability, meaning that the suggested approximation would be inaccurate and coding of the full variance (or, perhaps application of an appropriate bootstrapping method) would be required.

The proposed methods require the consistency of the IPCW weight. Therefore, the proportional hazards model for dependent censoring should be correctly specified. This may be approximately true when a sufficient number of both baseline and time-dependent covariates are incorporated.

In our analysis of chronic liver disease patients wait-listed for liver transplantation, we found that (calculated) Model for End-stage Liver Disease scores of 14–15 were consistent with the pre-transplant mortality hazard of patients granted an exception due to hepatocellular cancer. The results based on case-control sampling were consistent with those from the full-cohort analysis. In contrast, if no adjustment was made for dependent censoring, it appeared that MELD 16–17 and HCC had equivalent pre-transplant mortality. Therefore, our results indicate that the current MELD exception score of 22 granted to HCC patients overstates the actual medical urgency, and that an assigned MELD score of 14 or 15 may be more appropriate. If this adjustment were actually implemented, it would considerably decrease the frequency with which HCC patients receive liver transplants. The results of our analysis provide currently the most definitive evidence that HCC patients are given inappropriately high priority for liver transplantation in the United States.

In certain settings, the factor of interest (in addition to the adjustment covariates) may be time-dependent. In our analysis of the liver disease data, HCC designations are typically made at time 0, implying that the appropriate comparison is to patients not classified as HCC at time 0. However, if we were interested in evaluating exception scores generally, then the most useful analysis which

addressed the appropriateness of (what would primarily be non-HCC) exception scores would probably use time-dependent factors in the death model. This is because a hepatologist may observe a patient clearly getting worse, but the patient's MELD score not increasing, implying that the score does not accurately reflect the patient's need for liver transplantation. This analysis is currently under our consideration, and it appears that the proposed methods cannot readily be extended to accommodate such settings.

APPENDIX

Proofs of Theorem 2 and Theorem 3 are provided in the Supplementary Materials document.

ACKNOWLEDGEMENTS

The authors thank the Associate Editor and Referees for their constructive comments and suggestions, which resulted in a substantially improved manuscript. This work was supported in part by National Institutes of Health Grant R01 DK070869. The authors also thank the Scientific Registry of Transplant Recipients (SRTR) access to the end-stage liver disease data. The SRTR is funded by a contract from the Health Resource and Service Administration, U.S. Department of Health and Human Services.

BIBLIOGRAPHY

Barlow, W. E. (1994). Robust variance estimation for the case-cohort design. *Biometrics*, 50, 1064–1072.

Basto, S. T., Perez, R., Villela-Nogueira, C., Fernandes, E., Barroso, A., Victor, L., Pereira, B., Ribeiro, J. & Coelho, H. (2008). MELD as a predictor of long term mortality in liver transplantation waiting list. *Liver Transplantation*, 14, S219–S219.

Binder, D. A. (1992). Fitting Cox's proportional hazards models from survey data. *Biometrika*, 79, 139–147.

Borgan, Ø., Langholz, B., Samuelsen, S. O., Goldstein, L. & Pogoda, J. (2000). Exposure stratified case-cohort designs. *Lifetime Data Analysis*, 6, 39–58.

Breslow, N. E. & Holubkov, R. (1997a). Maximum likelihood estimation of logistic regression parameters under two-phase outcome-dependent sampling. *Journal of the Royal Statistical Society Series B*, 59, 447–461.

Chen, K. (2001). Generalized case-cohort sampling. *Journal of the Royal Statistical Society Series B*, 63, 791–809.

Chen, K. & Lo, S. H. (1999). Case-cohort and case-control analysis with Cox's model. *Biometrika*, 86, 755–764.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society Series B*, 34, 187–220.

Gorfine, M., Zucker, D. M., & Hsu, L. (2009). Case-control survival analysis with a general semiparametric shared frailty model: A pseudo full likelihood approach. *The Annals of Statistics*, 37, 1489–1517.

Gray, R. J. (2009). Weighted analyses for cohort sampling designs. *Lifetime Data Analysis*, 15, 24–40.

Hernán, M. A., Brumback, B., & Robins, J. M. (2000). Marginal structural models to estimate the causal effect of Zidovudine on the survival of HIV-positive men. *Epidemiology*, 11, 561–570.

Hsu, L. & Gorfine, M. (2006). Multivariate survival analysis for case-control family data. *Biostatistics*, 7, 387–398.

Hsu, L., Chen, L., Gorfine, M. & Malone, K. (2004). Semiparametric estimation of marginal hazard function from case-control family studies. *Biometrics*, 60, 936–944.

Huo, T. I., Wu, J. C., Lin, H. C., Lee, F. Y., Hou, M. C., Lee, P. C., Chang, F. Y. & Lee, S. D. (2005). Evaluation of the increase in model for end-stage liver disease (MELD) score over time as a prognostic predictor in patients with advanced cirrhosis: risk factor analysis and comparison with initial MELD and Child-Turcotte-Pugh score. *Journal of Hepatology*, 42, 826–832.

Kalbfleisch, J. D. & Lawless, J. F. (1988). Likelihood analysis of multi-state models for disease incidence and mortality. *Statistics in Medicine*, 7, 149–160.

Kalbfleisch, J. D. & Prentice, R. L. (2002). *The statistical analysis of failure time data*, 2nd ed., Wiley, Hoboken, NJ.

Kremers, W. K., van IJperen, M., Kim, W. R., Freeman, R. B., Harper, A. M., Kamath, P. S. & Wiesner, R. H. (2004). MELD score as a predictor of pretransplant and posttransplant survival in OPTN/UNOS status 1 patients. *Hepatology*, 39, 764–769.

Kulich, M. & Lin, D. Y. (2000). Additive hazards regression for case-cohort studies. *Biometrika*, 87, 73–87.

Li, H., Yang, P., & Schwartz, A. G. (1998). Analysis of age at onset from case-control family studies. *Biometrics*, 54, 1030–1039.

Lin, D. Y. (2000). On fitting Cox's proportional hazards models to survey data. *Biometrika*, 87, 37–47.

Lin, D. Y. & Ying, Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association*, 88, 1341–1349.

Merion, R. M., Schaubel, D. E., Dykstra, D. M., Freeman, R. B., Port, F. K. & Wolfe, R. A. (2005). The survival benefit of liver transplantation. *American Journal of Transplantation*, 5, 307–313.

Moger, T. A., Pawitan, A. and Borgan, Ø. (2008). Casde-cohort methods for survival data on families from routine registers. *Statistics in Medicine*, 27, 1062–1074.

Nan, B., Kalbfleisch, J. D., & Yu, M. (2009). Asymptotic theory for the semiparametric accelerated failure time model with missing data. *Annals of Statistics*, 37, 2351–2376.

Oakes, D. (1981). Survival times: aspects of partial likelihood (with discussion). *International Statistical Review*, 49, 235–264.

Pan, Q. & Schaubel, D. E. (2008). Proportional hazards models based on biased samples and estimated selection probabilities. *Canadian Journal of Statistics*, 36(1), 111–127.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, 73, 1–11.

Prentice, R. L. & Brewlow, N. E. (1978). Retrospective studies and failure time models. *Biometrika*, 65, 153–158.

Robins, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association*, 24–33.

Robins, J. M. & Finkelstein, D. M. (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*, 56, 779–788.

Robins, J. M. & Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology—Methodological Issues*, Jewell N. P., Dietz K., & Farewell V. T., editors. Birkhäuser, Boston, pp. 297–331.

Rubin, D. B. (1977). Inference and missing data. *Biometrika*, 63, 581–592.

Schaubel, D. E., Wolfe, R. A., Sima, C. S. & Merion, R. M. (2009). Estimating the effect of a time-dependent treatment by levels of an internal time-dependent covariate. *Journal of the American Statistical Association*, 104, 49–59.

Schildcrout, J. S. & Heagerty, P. J. (2008). On outcome-dependent sampling designs for longitudinal binary response data with time-varying covariates. *Biostatistics*, 9, 735–749.

Self, S. G. & Prentice, R. L. (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. *Annals of Statistics*, 16, 64–81.

Shih, J. H. & Chatterjee, N. (2002). Analysis of survival data from case-control family studies. *Biometrics*, 58, 502–509.

Song, R., Zhou, H., & Kosorok, M. R. (2009). On semiparametric efficient inference for two-stage outcome-dependent sampling with a continuous outcome. *Biometrika*, 96, 221–228.

Subramanian, A., Sulkowski, M., Barin, B., Stablein, D., Curry, M., Nissen, N., Dove, L., Roland, M., Florman, S., Blumberg, E., Stosor, V., Jayaweera, D. T., Huprikar, S., Fung, J., Pruett, T., Stock, P. &

Ragni, M. (2010). MELD score is an important predictor of pretransplantation mortality in HIV-infected liver transplant candidates. *Gastroenterology*, 138, 159–164.

Wang, W., Scharfstein, D., Tan, Z. & MacKenzie, E. J. (2009). Causal inference in outcome-dependent two-phase sampling designs. *Journal of the Royal Statistical Society Series B*, 71, 947–969.

Wiesner, R. H., McDiarmid, S. V., Kamath, P. S., Edwards, E. B., Malinchoc, M., Kremers, W. K., Krom, R. A. & Kim, W. R. (2001). MELD and PELD: Application of survival models to liver allocation. *Liver Transplantation*, 7, 567–580.

Zhou, H., Weaver, M. A., Qin, J., Longnecker, M. P. & Wang, M. C. (2002). A semiparametric empirical likelihood method for data from an outcome-dependent sampling scheme with a continuous outcome. *Biometrics*, 58, 413–421.

Received 23 December 2012
Accepted 31 March 2014