Locally Efficient Estimation of the Quality Adjusted Lifetime Distribution with Right-Censored Data and Covariates.

Mark J. van der Laan and Alan Hubbard

Division of Biostatistics

University of California

Berkeley, CA 94720

May 15, 2001

Abstract

Zhao and Tsiatis (1997) consider the problem of estimation of the distribution of the quality adjusted lifetime when the chronological survival time is subject to right-censoring. The quality adjusted lifetime is typically defined as a weighted sum of the times spend in certain states up till death or some other failur e time. They propose an estimator and establish the relevant asymptotics, under the assumption of independent censoring. In this paper we extend the data structure with a covariate process observed till end of follow up and identify the optimal estimation problem. Because of the curse of dimensionality, no globally efficient nonparametric estimators with a good practical p erformance at moderate sample sizes exist. Given a correctly specified model for the hazard of censoring conditional on the observed state process, we propose a closed form one-step estimator of the distribution of the quality adjusted lifetime whose asymptotic variance attains the efficiency boun d, if we can correctly specify a lower-dimensional working model for the conditional distribution of quality adjusted lifetime given the observed quality of life and covariate processes. The estimator remains consistent and asymptotically normal even if t his latter submodel is misspecified. The practical performance of the estimators is illustrated with a simulation study.

We also extend our proposed one-step estimator to the case treatment assignment is confounded by observed risk factors so that it can be used to test a treatment effect in an observational study.

Some key words: Asymptotically efficient, Asymptotically linear estimator, Cox proportional hazards model, Influence curve, Right censored data.

1 Introduction.

Gelber, et al. (1991) discuss quality of life data concerning operable breast cancer. The study compares a single cycle of adjuvant chemotherapy compared with longer duration chemotherapy for premenopausal women or chemoendrocrine therapy for postmenopa usal women. To define the quality adjusted lifetime, they considered three health states; time without systems and toxicity (TWiST), toxicity (TOX) and survival time after relapse (REL). They weighted the time spent in each category according to subject ive judgements as to the quality of life in each state. Specifically, TWiST was weighted as 1 with the other states naturally having less weight. Their goal was to compare the efficacy of different treatment regimes on the weighted sum of the times, kno wn as the quality adjusted lifetime and the parameter of interest to measure this was the mean quality adjusted lifetime in each treatment group. In this paper we present estimators of the distribution of quality adjusted lifetime and estimates of their variance. In addition, we present estimators that can estimate treatment specific quality adjusted lifetime distributions even when treatments are not assigned completely at random. These estimators and their standard errors can then be used to compare treatment efficacy on the quality adjusted lifetime distribution

In addition to Gelber, et al. (1991), many statistical analyses of lifetime data from clinical trials are concerned with inference on the quality adjusted lifetime distribution (e.g. Gelber, et al., 1989, Glasziou, et al., 1990 and Korn, 1993). In the se studies, one observes for each subject a quality of life state process up till the minimum of chronological death and censoring and weights are assigned to each of the states given a priori (an overview of this type of data is given by Cox et al., 199 2). Recently, Zhao and Tsiatis (1997) proposed an estimator of the quality adjusted lifetime distribution itself and they also provide a consistent estimator of its asymptotic variance. We will extend the work of Zhao and Tsiatis (1997). First, using g eneral results in Robins (1993) and Robins and Rotnitzky (1992), we identify the optimal semiparametric information bound in their model and we construct estimators achieving these optimal bounds. Moreover, their estimator is inconsistent if censoring, i .e. the follow up time, depends on the quality adjusted lifetime through the observed quality of life process. Dependent censoring occurs when the subject's condition affects the censoring time. For example, the decision to change treatment, and thus t o censor the subject, can be based on the quality of life and covariate processes of the subject. Therefore it remains to construct estimators that allow censoring to be statistically dependent upon on these observed processes. In addition, one frequen tly will also observe for each subject a covariate process (time-independent and time-dependent covariates). In this paper we will provide an estimator that 1) can incorporate additional time-dependent and time-independent covariates 2) is (locally) effi cient and 3) allows for dependent censoring. In addition, we provide an important modification of this estimator that can estimate a treatment-specific distribution of quality adjusted lifetime even if two or more treatments have not been assigned to th e population completely at random. Specifically, our estimators provide consistent estimates of the treatment distributions if the effect of treatment is confounded by time-independent covariates.

1.1 The data structure.

Let T be the chronological survival time of the subject. Let $V(t)$ be the state (condition) of the subject at time t. Typically, one assumes that the state space is finite so that $V(t) \in \{1, \ldots, k\}$ for some k, but this is not necessary. As in Zhao and Tsiatis (1997) we define the quality adjusted lifetime as $U \equiv \int_0^T Q(V(t))dt$, where Q is some given function. We remark here that the estimators proposed in this paper apply to any U which is a known function of the state process $\{V(t) : t \in [0, T]\}$. One sensible definition of quality adjusted lifetime can be obtained by defining $V(t)$ as the quality of life at time t on a scale between 0 and 1 and $Q(t) = t$. In this case, one might define $V(t) = 1$ if the subject has a normal life at time t and $V(t) = 0$ if the subject is seriously suffering from the treatment (e.g. chemotherapy) at time t.

In many applications one will observe additional information on the subject in terms of baseline covariates and time-dependent measurements (e.g, CD4-count in an AIDS-application). Let $W(t) \in \mathbb{R}^k$, $t \in \mathbb{R}_{\geq 0}$ be a covariate process. We will denote the time-independent covariates with $W(0)$. Now, the full-data process of interest is $X(t) = (R(t), V(t), W(t))$, where $R(t) = I(T \leq t)$. For a time-dependent process X we define $\bar{X}(t) = \{X(s) : s \leq t\}$ as the sample path of X up till time t. The full data of interest is $X \equiv \bar{X}(T) = (T, \bar{V}(T), \bar{W}(T))$. Note that observing X implies observing U. Therefore, if each subject is observed till death then the natural nonparametric estimator of the quality adjusted survival function $S_U(t) \equiv P(U > t) = 1 - F_U(t)$ is the empircal survival function based on $U_1, \ldots, U_n$. Due to limited follow up time or other reasons, one will not always observe the process X up until time T ; one observes the process X up until the minimum of a censoring time C and T and one knows whether this minimum is either the censoring time or T. Thus the observed data structure can be represented as:

$$Y = (\tilde{T} \equiv C \wedge T, \Delta = I(\tilde{T} = T), \bar{X}(\tilde{T})) \tag{1}$$

We observe n independent and identically distributed observations $Y_1, \ldots, Y_n$ of Y.

Robins (1993) and Robins and Rotnitzky (1992) proposed locally efficient estimators of the distribution of T based on n observations of $(T \wedge C, \Delta = I(T \leq C), \bar{L}(T \wedge C))$ for some process L related to T. Since we succeeded in rep resenting the data structure in the same fashion, we will be able to apply their theoretical results for construction of closed form locally efficient estimators of $F_U(t)$.

1.2 The model.

In this section, we describe the model for Y. Note that Y is a function of X and C and thus its distribution is indexed by the distribution of X and the conditional distribution of C, given X. The distribution F_X of X will be completely unspecified and we will assume that the conditional distribution $G(\cdot \mid X)$ of C, given X, satisfies "coarsening at random" (CAR). In words, the censoring mechanism satisfies CAR if censoring is not informative, given the observed covariates, as was originally formulated in Heitjan and Rubin (1991) and generalized in Jacobsen and Keiding (1995) and Gill,et al. (1997). It follows that $G(\cdot \mid X)$ satisfies CAR if for $c < T$

$$\lambda_C(c \mid X) = m(c, \bar{X}(c)) \text{ for some function } m \text{ of } (c, \bar{X}(c)), \tag{2}$$

where $\lambda_C(c \mid X)$ is the Lebesgue hazard corresponding with $G(dc \mid X)$ (Robins, 1993). The importance of the CAR assumption in estimation of parameters of F_X in the presence of a time-dependent surrogate process has been argued by Robins an d Rotnitzky (1992) who illustrate that, in many applications, the probability of censoring in $(t, t+\delta)$ depends on the observed covariate history up until time t. For example, a physician decides to change treatment at time t, where his decision is based on the observed surrogate process W and state process V up until time t; this type of dependent censoring satisfies (2).

By the curse of dimensionality, asymptotically efficient estimators, like a smoothed non-parametric maximum likelihood estimator, perform poorly in this context. Even without covariate process W the data structure still includes a time-dependent process V, which makes any fully efficient estimator impractical. In Gill, et al. (1997) it is shown that if (2) is the only assumption, then the model is saturated so that every regular and asymptotically linear estimator of $S_U(t)$ is asymptotically equivalent and thus efficient. Therefore, it is only possible to construct sensible estimators if one makes a stronger assumption than (2). Zhao and Tsiatis (1997) succeeded in constructing an ad hoc estimator by assuming that C is independe nt of X (which equals $(T, \bar{V}(T))$ in their marginal model).

To construct our estimators, we assume a parametric or semiparametric submodel of (2) for $G(\cdot \mid X)$. Let

$$\lambda_C(c \mid X) = m_\eta(c, \bar{X}(c)) \text{ for some model } m_\eta,\ \eta \in \Gamma. \tag{3}$$

In this paper, we will emphasize modeling λ_C with the Cox proportional hazards model using summary measures of $\bar{X}(C)$ as time-dependent and time-independent covariates:

$$\lambda_C(c \mid x) = \lambda_0(c) \exp\left(\alpha_0^\top W_1(c)\right), \tag{4}$$

where $\alpha_0 \in \mathbb{R}^k$ and $W_1(c) = f(\bar{X}(c)) \in \mathbb{R}^k$ is a vector of known functions of $\bar{X}(c)$. Thus one can include time-dependent covariates extracted from the state process $\bar{V}(c)$ and the covariate process $\bar{W}(c)$. This Cox proportional hazards model for censoring includes the independent censoring model of Zhao and Tsiatis (1997).

For a given realization of V we define

$$T(t,V) \equiv \inf\left\{s : \int_0^s Q(V(s))ds \geq t\right\} \tag{5}$$

as the time-point at which the subject has accumulated quality adjusted lifetime equal to t. Our results for $F_U(t)$ require that:

$$\bar{G}(T(t,V) \mid X) > 0 \ F_X\text{-a.e.} \tag{6}$$

1.3 Organization of paper.

In section 2, following the terminology of Robins (1993) and Robins and Rotnitsky (1992), we introduce an "inverse probability of censoring weighted" (IPCW) estimator of $S_U(t)$. In section 3, we will define a locally efficient one-step estimator in te rms of our IPCW-estimator plus an empirical mean of an estimator of the efficient influence curve of $F_U(t)$. This efficient influence curve is a function of the conditional probability $F(t \mid \bar{X}(u), \tilde{T} > u) = P(U \leq t \mid \bar{X}(u), \tilde{T} > u)$, and we provide a generic estimation method for estimating this conditional distribution. In addition, we show that we obtain without any additional work a confidence interval which has the right-coverage if $F(t \mid \bar{X}(u), \tilde{T} > u)$ is estimated consistently and is conservative otherwise. In section 4, we generalize our proposed one-step estimator to the case of non-random treatment assignment, that is, to the situation when treatment is confounded by observed risk factors. Thus, our one-step estimator can be used to estimate the distribution of the quality adjusted survival time U_A (U_B), where U_A (U_B) is defined as the quality adjusted survival time if the subject receives treatment A (B). From these estimate s and estimates of their asymptotic variances, one can construct a nonparametric locally efficient test of a treatment effect on survival in an observational study. Finally, in section 5 we present a simulation study comparing the estimator proposed by Zhao and Tsiatis (1997) with the IPCW-estimator and locally efficient one-step estimator.

In this paper, we do not discuss in detail the theoretical underpinnings of our estimators but defer to Hubbard, et al. (1998) who have summarized the theory of Robins and Rotnitzky (1992) and Robins (1993) as applied to nonparametric estimation of surviv al with right-censored data. Because the data and estimation problem we discuss in this paper are equivalent to that discussed in Hubbard et al. (1998), their results can be applied directly to estimation of the quality adjusted lifetime distribution. T he theory discussed in that paper states that,assuming the "regularity" conditions to be true and a correctly specified model for the censoring mechanism, then our one-step estimator is consistent and asymptotically normal. In addition, if one estimat es the conditional distribution $F(\cdot \mid \bar{X}(u), \tilde{T} > u)$ consistently, then the resulting one-step estimator of $F_U(t)$ is efficient. This allows one to guess a working model for the conditional distribution of U, given $\bar{X}(u),\ \ tildeT > u$. As a result, the one-step estimator is efficient if the working model contains the true $F(\cdot \mid \bar{X}(u), \tilde{T} > u)$ and remains consistent and asymptotically normal if it is misspecified, that is, the estimator is *locally* e fficient at the working model.

2 Inverse probability of censoring weighted estimators

In this section, we define an initial estimator for our one-step estimator that weights the observed $I(U_i \leq t)$ by the correct probability of censoring. We will call this estimator the "Inverse probability of censoring weighted estimator" (IPCW) estimator, following the terminology of Robins and Rotnitzky (1992). We exploit the following key identity to construct this IPCW-estimator (given (6)):

$$E\left\{\frac{I(U \leq t)\Delta}{\bar{G}(\tilde{T} \mid X)}\right\} = F_U(t), \tag{7}$$

where $\bar{G}(c \mid X) = P(C \geq c \mid X)$ denotes the conditional survival function of C, given X. This identity follows directly from

$$E(\Delta \mid X) = P(C \geq T \mid X) = \bar{G}(T \mid X),$$

which shows that the conditional expectation given X of the left-hand side of (7) equals $I(U \leq t)$. This suggest the following ad hoc estimator of $F_U(t)$:

$$F_n^0(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{I(U_i \leq t)\Delta_i}{\bar{G}_n(\tilde{T}_i \mid X_i)}. \tag{8}$$

where $\bar{G}_n$ is an estimator of $\bar{G}$ assuming the given model (3). Note that by the coarsening at random assumption (2), $\bar{G}(\tilde{T} \mid X)$ is only a function of $Y = (\tilde{T}, \Delta = I(\tilde{T} = T), \bar{X}(\tilde{T}))$ so that $F_n^0(t)$ indeed only depends on $Y_1, \ldots, Y_n$. If one assumes Cox proportional hazard for G, then one can use standard software to obtain the maximum (partial) likelihood estimator of the baseline hazard and the regression coefficients. I f one assumes that C is completely independent of X as in Zhao and Tsiatis (1997), then one can consistently estimate $\bar{G}$ with the Kaplan-Meier estimator based on the n observations of $(\tilde{T}, \Delta)$, where now T plays the role of the censoring variable. In the case that $G(\cdot \mid X) = G(\cdot)$ and $\bar{G}_n$ is the Kaplan-Meier estimator, simulations in section 5 show that this estimator (8) is competitive with the more complicated estimator proposed in Zhao and Tsiatis (1997).

3 The locally efficient one-step estimator.

In this section, we construct a locally efficient one-step estimator by adding to the estimator in $F_n^0(t)$ (8) an estimate of the empirical mean of the efficient influence function. In order to construct this one-step estimator, one first need s a representation of this efficient influence function. Robins and Rotnitsky (1992) and Robins (1993) presented the efficient influence function for estimation with right-censored data and thus we can apply their results to the data structure outlined in section 1.1 (in addition, see Hubbard, et al., 1998). Our representation of the efficient influence curve has two pieces. The first is given by the influence function of $F_n^0(t)$ using the known G, which is given by:

$$IC_0(Y \mid G, F_U(t)) \equiv \frac{I(U \leq t)\Delta}{\bar{G}(\tilde{T} \mid X)} - F_U(t). \tag{9}$$

The second piece, which is a function of the observed data Y, is defined by:

$$IC_{nu}^*(Y \mid F_X, G) = -\int_0^{T(t,V)} F(t \mid \bar{X}(u), \tilde{T} > u) \frac{dM(u)}{\bar{G}(u \mid X)}, \tag{10}$$

where

$$dM(u) \equiv I(C \in du, \Delta = 0) - \Lambda_C(du \mid X) I(\tilde{T} > u) \tag{11}$$

and $F(t \mid \bar{X}(u), \tilde{T} > u)$ is the conditional probability that $U \leq t$, given $\bar{X}(u)$ and $\tilde{T} > u$.

By results summared in Hubbard et al. (1993), the efficient influence curve for estimation of $F_U(t)$ is

$$IC^*(Y \mid F_X, G, F_U(t)) \equiv IC_0(Y \mid G, F_U(t)) - IC_{nu}^*(Y \mid F_X, G). \tag{12}$$

Following Robins and Rotnitzky (1992) we will slighly adjust this representation of the efficient influence curve. This adjustment will typically result in a slightly more efficient one-step estimator when the model for F_X is misspecified. In the following we use the notation $IC_{nu}^*(Y \mid F_X^1, G)$ for the expression (10) with the true conditional distribution $F(t \mid \bar{X}(u), \tilde{T} > u)$ replaced by a possibly incorrect $F^1(t \mid \bar{X}(u), \tilde{T} > u)$. We define for a given F_X^1, $G, F_U(t)$:

$$IC^*(Y \mid F_X^1, G, F_U(t)) \equiv IC_0(Y \mid G, F_U(t)) - c(F_X^1, G, F_U(t)) IC_{nu}^*(Y \mid F_X^1, G), \tag{13}$$

where

$$c(F_X^1, G, F_U(t)) \equiv \frac{E\left\{IC_0(Y \mid G, F_U(t)) IC_{nu}^*(Y \mid F_X^1, G)\right\}}{E\left\{(IC_{nu}^*(Y \mid F_X^1, G))^2\right\}}.$$

Here the expectation is always taken w.r.t. the true distribution of Y. This constant only has an effect at misspecification, that is $c(F_X, G, F(t)) = 1$, whereas $c(F_X^1, G, F(t)) \neq 1$ for $F_X^1 \neq F_X$ (Hubbard et al., 1998).

Since for any function f

$$\text{var}\left\{IC_0(Y \mid G, F_U(t)) - f(Y)\right\} \geq \text{var}\left(IC_0(Y \mid G, F_U(t)) - \frac{E(IC_0(Y)f(Y))}{Ef^2(Y))} f(Y)\right),$$

with equality if and only if the projection of IC_0 on f equals f, where f plays the role of $IC_{nu}^*(\cdot \mid F_X^1, G)$ (which represents the limit of our estimator $IC_{nu}^*(\cdot \mid F_{X,n}, G_n)$). Thus, the adjusted efficient infl uence curve (13) at $(F_X^1, G, F_U(t))$ has a smaller (or equal, if $F_X^1 = F_X$) variance than the unadjusted efficient influence curve (12) at $(F_X^1, G, F_U(t))$.

Let $IC_{nu}^*(\cdot \mid F_{X,n}, G_n)$ be an estimator of $IC_{nu}^*(\cdot \mid F_X, G)$ obtained by substitution of estimators of $F(t \mid \bar{X}(u), \tilde{T} > u)$ and G. Note that IC_{nu}^* depends on G also through the measure $dM(u)$. I n the next subsection we propose an estimator of $F(t \mid \bar{X}(u), \tilde{T} > u)$. One can estimate $c(F_X, G, F_U(t))$ with

$$c_n = \frac{\sum_{i=1}^n IC_0(Y_i \mid G_n, F_n^0(t)) IC_{nu}^*(Y_i \mid F_{X,n}, G_n)}{\sum_{i=1}^n \{IC_{nu}^*(Y_i \mid F_{X,n}, G_n)\}^2}, \tag{14}$$

where $F_{X,n}$ denotes the estimator of $F(t \mid \bar{X}(u), \tilde{T} > u)$. Now, we estimate the efficient influence curve IC^* by substituting these estimators in the representation (13):

$$IC^*(Y \mid F_{X,n}, G_n, F_n^0(t)) = IC_0(Y \mid G_n, F_n^0(t)) - c_n IC_{nu}^*(Y \mid F_{X,n}, G_n), \tag{15}$$

where F_n^0 is the IPCW-estimator defined in (8).

We propose to estimate $F_U(t)$ with the one step estimator:

$$F_n^1(t) = F_n^0(t) + \frac{1}{n}\sum_{i=1}^n IC^*(Y_i \mid F_{X,n}, G_n, F_n^0(t)) \tag{16}$$

$$= F_n^0(t) + \frac{1}{n}\sum_{i=1}^n \left\{IC_0(Y_i \mid G_n, F_n^0(t)) - c_n IC_{nu}^*(Y_i \mid F_{X,n}, G_n)\right\}. \tag{17}$$

One could also just set $c_n = 1$ and still obtain a locally efficient estimator. However, if $F_n(t \mid \bar{X}(u), \tilde{T} > u)$ is inconsistent, then the estimator using the empirical c_n is typically asymptotically more efficient than the estimator using $c_n = 1$, whereas if $F_n(t \mid \bar{W}(u), \tilde{T} > u)$ is consistent, then $c_n \rightarrow 1$. Note that $F_n^1(t)$ is just the classical one-step estimator as defined in Bickel et al. (1993, page 395).

The one-step estimator $F_n^1(t)$ depends on estimates of G and $F(t \mid \bar{X}(u), \tilde{T} > u)$. Theorem 5.1 in Hubbard et al. (1998) can be used as a template to prove the local efficiency result for the one-step estimator $F_n^1(t)$. Generally spe aking, this theorem shows that if G_n is estimated consistently, then the estimator $F_n^1(t)$ is consistent and asymptotically linear and if IC_{nu}^*, i.e. $F(t \mid \bar{X}(u), \tilde{T} > u)$, is also estimated consistently, then $F_n^1(t)$ is ev en asymptotically efficient. In addition, the theorem shows that $F_n^1(t)$ is asymptotically linear with an influence curve which has variance smaller than the variance of the limit $IC^*(Y \mid F_X^1, G, F_U(t))$ of the estimator $IC^*(Y \mid F_{X,n}, G_n, F_n^0(t))$ used in the one-step estimator. In particular, this shows that the variance, even under complete misspecification of $F(t \mid \bar{X}(u), \tilde{T} > u)$, is still smaller than the variance of IC_0. Thus, by choosing IC_0 in our representation of the efficient influence curve equal to the influence curve of a given ad hoc estimator we can guarantee the one-step estimator to be more efficient than the ad hoc estimator. For example, we could have chosen IC_0 to be equal to the influence curve of the Zhao-Tsiatis estimator and obtain in this way an estimator which is even under complete misspecification more efficient than the Zhao-Tsiatis estimator. We choose not to complicate the representation of the one-step estimator since it is already locally efficient and this kind of modificati on is only practically useful under severe misspecification (see simulation section).

3.1 Estimation of IC^*_{nu} using a conditional expectation representation.

The idea of representing the conditional probability $F(t \mid \bar{X}(u), \tilde{T} > u)$ as a regression of a random variable $O_G(Y)$ on observed covariates is due to Robins (1993) and Robins and Rotnitzky (1992). It has a powerful application in estim ating $F(t \mid \bar{X}(u), \tilde{T} > u)$. First, we define the random variable

$$O_G \equiv \frac{I(U \leq t)\Delta \bar{G}(u \mid X)}{\bar{G}(\tilde{T} \mid X)}. \tag{18}$$

Under the assumption CAR (2) we have

$$F(t \mid \bar{X}(u), \tilde{T} > u) = E\left(O_G \mid \bar{X}(u), \tilde{T} > u\right). \tag{19}$$

Assume $\lambda_C(\cdot \mid x)$ satisfies the Cox-proportional hazards model (4). We can estimate α_0 with the partial likelihood equations only involving α and the corresponding estimator of the baseline-hazard, which itself is a simple function of this partial likelihood estimator and the data (Andersen, et al., 1993). This yields an estimator of M and $\bar{G}(u \mid X)$. Notice also that $d\hat{M}$ is only non-zero at the observed censoring times and thus if one uses the repr esentation (19), then it remains to estimate the conditional expectation of a random variable O_{G_n}, given $\bar{X}(u), \tilde{T} > u$, at a u corresponding with an observed censoring variable. Given an estimate G_n of G and for a given t, O_{G_n} is an observed random variable. Consequently, for every u corresponding with an observed C_i, one can carry out a parametric or nonparametric regression estimation of O_{G_n} on one or a number of relevant (for U) summary measures $Z_1, \ldots, Z_k$ of $\bar{X}(u)$, only using th e observations with $\tilde{T} > u$. In particular, we can assume that the regression is nonparametric additive and use the SPLUS function gam (Chambers and Hastie, 1993).

3.2 Construction of conservative confidence intervals.

Consider the one-step estimator (16) which is given by $F_n^0(t) + 1/n \sum_{i=1}^n \hat{IC}(Y_i)$, where $\hat{IC}(Y) \equiv IC^*(Y \mid F_{X,n}, G_n, F_n^0(t))$ defined by (15) then $F_n^1(t)$ is asymptotically linear with influence curve having variance smaller than or equal to the variance of $IC^*(Y \mid F_X^1, G, F_U(t))$, where $IC^*(Y \mid F_X^1, G, F_U(t))$ represents the limit for $n \to \infty$ of $\hat{IC}(Y)$ (Hubbard et al., 1998). Therefore a conservative estimate of the asymptotic variance of $F_n^1(t)$ is given by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\{\hat{IC}(Y_i)\}^2.$$

This can be used to construct a conservative 95% confidence interval for $F_U(t)$:

$$F_n^1(t) \pm 1.96\frac{\hat{\sigma}}{\sqrt{n}}. \qquad (20)$$

This confidence interval is conservative if one inconsistently estimates $F(t \mid \bar{X}(u), \tilde{T} > u)$ and it is asymptotically correct otherwise. This confidence interval is very practical since one gets if for free after having computed the estimator $F_n^1(t)$. Simulations in Hubbard, et al. (1998) suggest that the confidence interval (20) can be very close to the correct confidence interval, even at a high degree of misspecification.

4 Testing a treatment effect on quality adjusted survival in an observational study

In this section, we discuss a modification of both our IPCW and one-step estimators that can adjust for confounding of treatment by observed time-independent variables that can occur in observational studies. For simplicity, we will present our estimators in the context of two treatments, A and B; extrapolating to more than two treatments is straightforward. We also note that the treatment can also play the role of a discretized risk factor. Subsequently, we show how the estimators and theory presented above can be directly applied to the setting of non-random treatment assignment. This methodology is also presented in Hubbard, et al. (1998) for general right-censored data.

4.1 Model and data structure.

Consider a particular subject. Let $T_A, V_A(\cdot), W_A(\cdot)$ be the chronological survival time, the state process and the covariate process, respectively, of the subject when receiving treatment A. Let $X_A(t) = (I(T_A \leq t), V_A(t), W_A(t))$ be the full data process on the subject if he/she would receive treatment A. If the subject actually receives treatment A, then we would observe:

$$Y_A = (\tilde{T}_A \equiv C_A \wedge T_A, \Delta_A = I(\tilde{T}_A = T_A), \bar{X}_A(\tilde{T}_A)). \tag{21}$$

The variable of interest is $U_A = \int_0^{T_A} Q(V_A(t))dt$. If everyone in our population was given treatment A, then we would observe n i.i.d copies $Y_{A,1}, \ldots, Y_{A,n}$ of Y_A. In that case we could apply our estimators as presented in section 2 and 3 above.

Similarly, we define Y_B as the data we would observe if the subject actually receives treatment B:

$$Y_B = (\tilde{T}_B \equiv C_B \wedge T_B, \Delta_B = I(\tilde{T}_B = T_B), \bar{X}_B(\tilde{T}_B)).$$

Note that $W_A(t)$ is a process that might contain both time-dependent and time-independent covariates. We will denote the time-independent covariates with $W(0)$ (measured at baseline $t = 0$). Thus $W(0) = X_A(0) = X_B(0)$ is a vector of covariates which is always observed.

In a statistically perfect word, we would observe, for each subject, both Y_A and Y_B. More typically subjects may be assigned either treatment A or B, so the data we observe has a missing data structure. The distribution of Y is indexed by th e distribution F_{X_s} of X_s, the distribution $G_s(C_s \mid X_s)$ of C_s given X_s and $P(S = s \mid X_A, X_B)$, $s \in \{A, B\}$. In our model $F_{s,X}$ will be completely unspecified and we will assume that the joint missingness distribution ($G_s(\cdot \mid X_s)$ and $P(S = s \mid X_A, X_B)$) is such that Y is "coarsening at random" (CAR) for (X_A, X_B). The missingness mechanism satisfies CAR, if for $c < T_s$,

$$\begin{aligned} P(S = A \mid X_A, X_B) &= P(S = A \mid W(0)) \\ P(S = B \mid X_A, X_B) &= 1 - P(S = A \mid W(0)) \\ \lambda_{C,s}(c \mid X_s) &= m_s(c, \bar{X}_s(c)) \text{ for some function } m_s \text{ of } (c, \bar{X}_s(c)), \end{aligned} \tag{22}$$

where $\lambda_{s,C}(c \mid X)$ is the Lebesgue hazard corresponding with $G_s(dc \mid X_s)$, $s \in \{A, B\}$ and $P(S = A \mid X_A, X_B)$ is the so-called propensity score of Rubin (1978). CAR implies that likelihood factorizes into a $F_{s,X}$ part, a G_s part and a $P(S = \cdot \mid X_A, X_B)$ part.

If one is only interested in the distribution of X_A, and one assumes no relationship between X_A and X_B (as we will do here), then X_A is the full data. Therefore, for construction of a locally efficient estimator of $F_{U_A}(t)$ it makes sense to redefine the data Y as follows. Let C_A^* be the censoring variable, such that $C_A^* = C_A$ if $S = A$ and $C_A^* = 0$ if $S \neq A$ where S is the random treatment assignment. Then, one observes:

$$Y = (\tilde{T}_A^* \equiv \min(C_A^*, T_A), \Delta_A^* \equiv I(T_A < C_A^*) = I(T_A < C_A)I(S = A), \bar{X}_A(\tilde{T}_A^*)). \tag{23}$$

Note that $\bar{X}_A(0) = W(0)$ is observed regardless of treatment assignment. The data defined in this manner is the same as presented in section 1 except that the censoring distribution

$$G_A^*(c \mid X_A) = P(S = A \mid X_A)G_A(c \mid X_A). \tag{24}$$

corresponding to the joint conditional distribution of censoring and treatment assignment, is now playing the role of G. To find the distribution of U_B, we can define the full data now as X_B and create a new censoring variable, C_B^*, which equ als C_B if $S = B$ and 0 otherwise.

To construct estimators with good (locally efficient) finite sample performance, we assume a parametric or semiparametric submodel of (22) for both $G_s(\cdot \mid X_S)$, $s \in \{A, B\}$, and $P(S = A \mid X_A)$. We can model $P(S = A \mid W(0))$ with a logistic linear or additive regression model and use standard software (e.g. Splus function *glm* or *gam*) to estimate the regression (Chambers and Hastie, 1993). Finally, our results for estimation of $F_{U_A}(t)$ require that:

$$P(S = A \mid W(0))\bar{G}_A(T_A(t, V_A) \mid X_A) > 0, \; F_{X_A} \text{ a.e.} \tag{25}$$

In the next subsection we propose locally efficient estimators of $F_{U_A}(t)$ (and similarly $F_{U_B}(t)$) under this model when observing n i.i.d. copies of Y.

4.2 Locally efficient estimators

Because the data structure is now identical to that presented in section 1.1, we can apply the same IPCW and one-step estimator as presented in sections 2 and 3: just replace G by G^*_A and C by C^*_A.

So define as in section 3

$$IC_0(Y \mid G^*_A, F_{U_A}(t)) = \frac{I(U_A \le t)\Delta^*_A}{\bar{G}^*(T_A \mid X_A)} - F_{U_A}(t) \tag{26}$$

$$IC^*_{nu}(Y \mid F_{X_A}, G^*_A) = -\int F_A(t \mid \bar{X}_A(u), \tilde{T}^*_A > u)\frac{dM^*_A(u)}{\bar{G}(u \mid X_A)}, \tag{27}$$

where

$$dM^*_A(u) = I(C^*_A \in du, \Delta^*_A = 0) - \Lambda_{C^*_A}(du \mid X_A)I(\tilde{T}^*_A > u).$$

Then the efficient influence curve for $F_{U_A}(t)$ is given by:

$$IC^*(Y \mid F_{X_A}, G^*_A, F_A(t)) = IC_0(Y \mid G^*_A, F_A(t)) - IC^*_{nu}(Y \mid F_{X_A}, G^*_A). \tag{28}$$

Note that

$$F_A(t \mid \bar{X}_A(u), \tilde{T}^*_A > u) = F_A(t \mid \bar{X}_A(u), \tilde{T}_A > u, S = A) \text{ for } u > 0$$

$$F_A(t \mid \bar{X}_A(0), \tilde{T}^*_A > 0) = F_A(t \mid W, S = A). \tag{29}$$

It is natural to parametrizes G^*_A as a product of G_A with $P(S = A \mid W(0))$ as in (24) and to estimate $F_A(t \mid \bar{W}_A(u), \tilde{T}^*_A > u)$ using the representation (29) in terms of $F_A(t \mid W, S = A)$ and $F_A(t \mid \bar{W}_A(u), \tilde{T}_A > u, S = A)$.

To implement the locally efficient estimator for $F_{U_A}(t)$, one must estimate the following components of the efficient influence curve IC^* (28): $P(S = A \mid W(0)), F_A(t \mid W(0), S = A), F_A(t \mid \bar{X}_A(u), \tilde{T}_A > u, S = A)$ and $G_A(u \mid X_A)$. Since S is independent of the X_A, given $W(0)$, we have:

$$P(C_A > t \mid X_A) = P(C_A > t \mid X_A, S = A).$$

Thus, one possible method of deriving estimates of $G_A(\cdot \mid X_A)$ and $F_A(t \mid \bar{X}_A(u), \tilde{T}_A > u, S = A)$ is to estimate them as discussed in section 3, using the subsample with $S = A$. Another possiblility is by using a traditional regres sion approach that includes a treatment dummy variable in the model.

Let $G^*_{A,n}$ the estimate of G^*_A obtained by plugging in estimators of G_A and $P(S = A \mid W(0))$ into (24). Thus, it remains to estimate $F_A(t \mid W(0), S = A)$. If C_A is independent of T_A, given $W(0)$, then one could assume, for example, a Cox-proportional hazards model with time-independent covariates $W(0)$ and estimate $F_A(t \mid W(0), S = A)$ with the partial likelihood estimator based on the sub-sample defined by $S = A$. In general, one can use the relation

$$F_A(t \mid W(0)) = E\left(\frac{I(T_A \le t)\Delta_A}{\bar{G}_A(T_A \mid X_A)} \mid W(0), S = A\right).$$

This can, of course, be estimated as a regression of $I(T_A \le t)\Delta_A / \bar{G}_{A,n}(T_A \mid X_A)$ on $W(0)$ using a flexible regression routine, such as the *gam*-function in Splus (Chambers and Hastie, 1993) on the sub-sample defined by $S = A$.

When all the components have been estimated, then we can represent our estimator, just as we did in section 3 (see 16), as a one-step estimator:

$$F_{A,n}^1(t) = F_{A,n}^0(t) + \frac{1}{n}\sum_{i=1}^{n} IC^*(Y_i \mid F_{X_A,n}, G_{A,n}^*, F_{A,n}^0(t)), \tag{30}$$

where

$$F_{A,n}^0(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{I(S_i = A)I(T_{A,i} \leq t)\Delta_i}{\bar{G}_{A,n}(T_{A,i} \mid X_{A,i})P_n(S_i = A \mid X_A)} \tag{31}$$

Here we choose, for the sake of simplicity, to set $c_n = 1$ although we could have estimated c_n using (14). Equivalent to our discussion in section 3, if $G_{A,n}$ and $P(S = A \mid W)$ are estimated consistently, then the estimator $F_{A,n}^1(t)$ is asymptotically linear. In addition, if $F_A(t \mid \bar{X}_A(u), \tilde{T}_A > u)$ and $F_A(t \mid W, S = A)$ are also estimated consistently, then $F_{A,n}^1(t)$ is asymptotically efficient. As in section 3 we estimate the influence curve IC_A of $F_{A,n}^1(t)$ with $IC_A^*(Y_i \mid G_{A,n}, P_n(S = A \mid W(0)), F_{X_A,n}, F_{A,n}^0(t))$. Similary, one can estimate the influence curve IC_B of $F_{B,n}^1$ with $IC_B^*(Y_i \mid G_{B,n}, P_n(S = B \mid W), F_{X_B,n}, F_{B,n}^0(t))$. Since $F_{A,n}^1(t) - F_{B,n}^1(t)$ is asymptotically linear with influence curve $IC_A - IC_B$ one can now estimate the limit variance of $F_{A,n}^1(t) - F_{B,n}^1(t)$ with

$$\hat{\sigma}^2 = 1/n \sum_i \Big\{ IC_A^*(Y_i \mid G_{A,n}, P_n(S = A \mid W), F_{X_A,n}, F_{A,n}^0(t)) - IC_B^*(Y_i \mid G_{B,n}, P_n(S = B \mid W), F_{X_B,n}, F_B^0$$

Thus a confidence interval for $F_A(t) - F_B(t)$ is given by:

$$F_{A,n}^1(t) - F_{B,n}^1(t) \pm 1.96\hat{\sigma}/\sqrt{n}.$$

This confidence interval is slightly conservative under misspecification and is correct otherwise. Of course, this confidence interval provides us also with a test for a treatment effect on survival at a specified t.

5 Simulation results.

We performed a simple simulation study to examine the relative performance of the competing estimators of the survival distribution: the biased Kaplan-Meier (S_{KM}), the IPCW-estimator

(S_{IPCW}), the Zhao-Tsiatis estimator (S_{ZT}), the locally ef ficient one-step estimator using only the process $\bar{V}(u)$ as covariate (S_{OS_1}), the one-step estimator using both $\bar{V}(u)$ and a relevant covariate W that uses a non-parametric smoother to estimate $F(t \mid \bar{X}(u), \tilde{T} > u)$ (S_{OS_2}) and the one-step estimator that uses an extreme parametric guess for estimating $F(t \mid \tilde{T} > u, \bar{X}(u))$ (S_{OS_3}). For this simulation, $W(0) = U + \sigma e$, $e \sim N(0, 1)$, $n = 100$ a nd the estimators are determined at each of 1000 random samples.

Our data generating model is similar to the first simulation in Zhao and Tsiatis (1997) and can by inspired by the data from Gelber, et al. (1991) discussed above. In that case, they defined three quality of life states: $TWiST$, TOX and REL. In thi s example, we will consider the quality adjusted time from treatment until relapse (REL) as the time variable of interest, U. Thus, before relapse there are two relevant states, TOX and $TWiST$, and for our simulations we will give TOX a weight of 0 and $TWiST$ a weight of 1. For our simulation, as in Gelber, et al. (1991), patients are administered the treatment at time 0 and this treatment is toxic and causes some period of illness for the patient. This period of toxicity, TOX, is gener ated randomly from $U(0, 100)$. Time to relapse from administration of treatment, REL, or in our notation, T, follows an exponential distribution with mean 120. Finally, some patients are not followed completely until relapse and are thus censored; c ensoring times, C, are generated from an exponential distribution with mean 100. Thus, one only knows the time spent without toxic side effects if one has observed the patient all the way from treatment to relapse. Thus $U = \max(0, T - TOX)$. To calculate S_{KM}, we use $C_{KM} \equiv \max(0, C - TOX)$ as the censoring time and $\Delta_{KM} \equiv I(U \leq C_{KM})$. As discussed above, this Kaplan-Meier estimator will be inconsistent because of the dependency between U and C_{KM}, through TOX.

The estimator S_{IPCW} is defined in (8). The Zhao and Tsiatis (1997) estimator, in our notation, is defined by:

$$S_{IPCW} = \frac{1}{n}\sum_{i=1}^{n}\frac{I(U_i > t)\Delta(t)}{\bar{G}_{n,ZT}(T_i(t,V_i))},$$

where $\Delta(t) = I(T(t,V) < C)$ and $\bar{G}_{n,ZT}$ is the Kaplan-Meier estimator of the censoring distribution where C is censored by $T(t,V)$. To calculate S_{OS_1}, one needs to decide how to incorporate the information $\bar{V}(u)$ into the e stimation of $F(t \mid \tilde{T} > u, \bar{V}(u))$. For this simulation, we use only the most recent value of the state process at time u, that is $V(u)$. To estimate $F(t \mid \tilde{T} > u, V(u))$ we first choose a set of u's that correspond to the observed censoring times. Then, for each u we use the expectation approach (19) where, for each u, we estimate $F(t \mid \tilde{T} > u, V(u))$ by taking a simple average of $O_{G_n,i}$ (18) over each of the two samples defined by $V(u) = 0$ and $V(u) = 1$, only including those observations with $\tilde{T} > u$.

Finally, we need to calculate the one-step estimators that use the covariate information. First of all, we use two values of σ that determine the strength of the covariate, $W(0)$, in predicting U: $\sigma = 0$ and $\sigma = 50$. In the first c ase, $W(0) = U$ and so if this information is utilized consistently, then $F_n(t \mid \tilde{T} > u, \bar{X}(u)) \to I(U < t)$ so that our one-step estimator should be asymptotically equivalent to the empirical distribution based on $U_1, \ldots, U_n$. In these simulations, S_{OS_2} and S_{OS_3} differ in the estimate of $F_n(t \mid \tilde{T} > u, \bar{X}(u))$. For the former, we use the nonparametric regression approach based on (19) as discussed in section 3.1. Specifically, we estimat e $F(t \mid \tilde{T} > u, V(u), W)$ by smoothly regressing the random variable O_{G_n} against $V(u)$ and W using a generalized additive models approach (Hastie and Tibshirani, 1990) and the super-smoother (Friedman, 1984) only on those observations with $\tilde{T} > u$. Note that this estimating model does not contain the truth if $\sigma = 0$. In contrast, S_{OS_3} uses an extreme parametric model for estimating F_X, that is $F_n(t \mid \tilde{T} > u, V(u) = v, W(0) = w) = I(w < t)$. When $\sigma = 0$, this should perform equivalently to the empirical based on the observed and unobserved U's. Thus, one can do no better. However, when $\sigma = 50$, then this

Table 1: Results of the simulation of time without toxicity. We report $MSE \times 10^2 (RMSE)$ for estimation of $F_U(t)$ at three quantiles.

Estimator	$t = 20$, $S_U(t) = 0.5$	$t = 65$, $S_U(t) = 0.4$	$t = 110$,$S_U(t) = 0.2$
S_{KM}	0.99(1.0)	0.85(1.0)	0.85(1.0)
S_{ZT}	0.45(2.2)	0.59(1.4)	0.72(1.2)
S_{IPCW}	0.43(2.3)	0.56(1.5)	0.69(1.2)
S_{OS_1}	0.43(2.3)	0.54(1.6)	0.68(1.2)
$\sigma = 0$			
S_{OS_2}	0.28(3.5)	0.36(2.3)	0.50(1.7)
S_{OS_3}	0.27(3.7)	0.25(3.4)	0.20(4.2)
$\sigma = 50$			
S_{OS_2}	0.36(2.7)	0.44(1.9)	0.51(1.7)
S_{OS_3}	0.45(2.2)	0.53(1.6)	0.55(1.5)

is a very inconsistent estimator of F_X, and one should expect poor utilization of the covariate information (although the estimator should remain consistent and asymptotically normal). These two met hods (smooth regression and an extreme parametric guess) were chosen to contrast the effects of using a low-dimensional model and highly non-parametric model when estimating $F(t \mid \tilde{T} > u, V(u) = v, W(0) = w)$.

The results of the simulation are shown in table 1. Note first the relatively poor performance of S_{KM}, due to the fact that S_{KM} is inconsistent. The simulation suggest that the performance of S_{ZT} and S_{IPCW} is nearly identical. We further investigated this by running a larger sample size ($n = 2000$) and found that their performance was still nearly identical, suggesting these estimators are asymptotically equivalent under the data-generating model used for this simulation. As expected, the one-step estimator using the extreme parametric guess (S_{OS_3}) does very well when the model is correct (when $\sigma = 0$). However, even the est imator using the more non-parametric regression model (S_{OS_2}) gains substantial efficiency relative to S_{IPCW} when $\sigma = 0$. When $\sigma = 50$, the "model" used in S_{OS_3} to estimate $F(t \mid \tilde{T} > u, \bar{V}(u), W(0))$ is now severely misspecified, and thus the performance deteriorates, but as we see, somewhat surprisingly, it still is as good as S_{IPCW}. Finally, because S_{OS_2} uses a flexible regression model to estimate F_X, it performs well when $\sigma = 50$. This occ urs despite the fact that now $W(0)$ contains only weak information about the location of U and the sample size is relatively small ($n = 100$) and the model used is not consistent.

For all of the one-step estimators in this simulation, we have set $c_n = 1$, in (16). If one uses extreme parametric models such as this, then one should use the adjustment with c_n and use for IC_0 the influence curve of the simple estimato r S_{IPCW}. In that case theory summarized in Hubbard, et al. (1998) proves that S_{0S_3} is asymptotically more efficient than S_{IPCW}, even when the guess is extremely wrong. However, this simulation shows that it takes a very poorly chosen model to make S_{0S_3} inferior to S_{IPCW}.

6 Conclusions

Zhao and Tsiatis (1997) proposed a sensible estimator that works well for estimating the marginal distribution of quality adjusted lifetime when C is independent of T and there are no relevant covariates. We proposed estimators that can account for dependent censoring and can utilize covariate information to increase the efficiency of estimation. In our simulation, the relative performance of our simple IPCW estimator is nearly equivalent to the estimator proposed by Zhao and Tsiatis. The simulation results also suggest that the gain in efficiency by using covariate information can be worthwhile even at relatively small sample sizes and that there is little risk in trying to use covariate information in the one-step estimator. If necessary, the one-step estimator can be adjusted (by using a IC_0 equal to the influence curve of a good ad hoc estimator) to be asymptotically more efficient than the chosen initial estimator. We also show how to construct locally efficient estimators of treatment distributions when treatment assignment is confounded by observed risk factors. The proposed estimators can be implemented using existing statistical software and give the analyst a powerful tool for dealing with this very interesting and practical type of data.

ACKNOWLEDGEMENT

This research was supported by a FIRST award from the National Institute of General Medical Sciences, National Institute of Health.

References

Andersen, P.K., Borgan, O., Gill, R.D. & Keiding, N. (1993), *Statistical models based on counting processes*, Springer, New York.

Bickel, P.J., Klaassen, A.J., Ritov, Y. & Wellner, J.A. (1993), *Efficient and adaptive inference in semi-parametric models*, Johns Hopkins University Press, Baltimore.

Cox,D.R., Fitzpatrick, R., Gore, E., Spiegelhalter, D.J. & Jones, D.R. (1992), Quality-of-life assessment: Can we keep it simple? *J. R. Statist. Soc.* **A 155**, 353–393.

Friedman, J. (1984), *A variable span smoother*, Department of Statistics, Technical report, Stanford University, Stanford, CA.

Gelber, R.D., Gelman, R.S. & Goldhirsch, A. (1989), A quality of life oriented endpoint for comparing therapies, *Biometrics* **45**, 781–795.

Gelber, R.D., Goldhirsch, A. & Cavalli, F. (1991), Quality-of-life-adjusted evaluation of a randomized trial comparing adjuvant therapies for operable breast cancer (for the international breast cancer study group), *Annals of Internal Med icine*, **114**, 621–628.

Gelber, R.D., Lenderking, W.R., Cotton D.J., Cole, B.F., Fischl, M., Goldhirsch, A. & Testa, M.A. (1992), Quality-of-life evaluation in a clinical trial of zidovudine therapy in patients with mildly symptomatic HIV infection, *Annals of Internal Medicine*, **116**, 961–966.

Glasziou, P. P., Simes, R.J. & Gelber, R.D. (1990), Quality adjusted survival analysis, *Statisti. Med.* **9**, 1259–1276.

Gill, R.D., van der Laan, M.J. & Robins, J.M. (1997), Coarsening at Random: Characterizations, Conjectures and Counter-Examples, *Proceedings of the First Seattle Symposium in Biostatistics*, 1995. D.Y. Lin and T.R. Fleming (editors), Springer Lecture Notes in Statistics, 255–294.

Goldhirsch, A., Gelber, R.D., Simes, R.J., Glasziou, P. P. & Coates, A. (1989), Costs and benefits of adjuvant therapy in breast cancer: a quality adjusted survival analysis., *J. Clin. Oncol.* **7**, 36–44.

Härdle, W. (1993), *Applied nonparametric regression*, Econometric Society Monographs, Cambridge University Press.

Hastie, T.J. & Tibshirani, R.J. (1990), *Generalized Additive Models*, Chapman and Hall, London.

Heitjan, D.F. & Rubin, D.B. (1991), Ignorability and coarse data, *Annals of Statistics* **19**, 2244–2253.

Jacobsen, M. & Keiding, N. (1995) , Coarsening at random in general sample spaces and random censoring in continuous time, *Ann. Statist.* **23**, 774–786.

Korn, E.L. (1993), On estimating the distribution function for quality of life in cancer clinical trials, *Biometrika* **80**, 535-542.

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. (1996), Locally efficient median regression with random censoring and surrogate markers, *Lifetime Data: Models in Reliability and Survival Analysis*, Ed. N.P. Jewell et al., Kluwer Academic Publishers, 263–274.

Robins, J.M. & Rotnitzky, A. (1992), "Recovery of information and adjustment for dependent censoring using surrogate markers," *Aids Epidemiology, Methodological issues*, Birkhäuser.

Rubin, D.B. (1978), Bayesian inference for causal Effects: The role of randomization, *Annals of Statistics*, **6**, 34–58.

van der Vaart, A.W. & Wellner, J.A. (1996), *Weak convergence and empirical processes*, Springer Verlag.

Zhao, H. & Tsiatis, A.A. (1997), A consistent estimator for the distribution of quality adjusted survival time, *Biometrika* **84**, 339–348.