FLEXIBLE HAZARD REGRESSION MODELING FOR MEDICAL COST DATA

BU-1547 -MA **April, 2001**

Robert L. Strawderman
Arvind K. Jain

Keywords: Cost-benefit analysis, Economic evaluation, Informative censoring; Quality of life.

Abstract:
The modeling of lifetime (i.e. cumulative) medical cost data in the presence of censored follow-up is complicated by induced informative censoring, rendering standard survival analysis tools invalid. With few exceptions, recently proposed nonparametric estimators for such data do not extend easily to handle covariate information. We propose to model the hazard function for lifetime cost endpoints using an adaptation of the HARE methodology (Kooperberg, Stone and Truong, *Journal of the American Statistical Association*, 1995, 90:78-94). Linear splines and their tensor products are used to adaptively build a model that incorporated covariates and covariate-by-cost interactions without restrictive parameter assumptions. The informative censoring problem is handled using inverse probability of censoring weighted estimating equations. The proposed method is illustrated using simulation and also with data on the cost of dialysis for patients with end-stage renal disease.

CORNELL UNIVERSITY BIOMETRICS DEPARTMENT TECHNICAL REPORT BU-1547-MA

Flexible Hazard Regression Modeling for Medical Cost Data

Robert L. Strawderman and Arvind K. Jain *

April 4, 2001

Abstract

The modeling of lifetime (i.e., cumulative) medical cost data in the presence of censored follow-up is complicated by induced informative censoring, rendering standard survival analysis tools invalid. With few exceptions, recently proposed nonparametric estimators for such data do not extend easily to handle covariate information. We propose to model the hazard function for lifetime cost endpoints using an adaptation of the HARE methodology (Kooperberg, Stone, and Truong, *Journal of the American Statistical Association*, 1995, **90**, 78-94). Linear splines and their tensor products are used to adaptively build a model that incorporates covariates and covariate-by-cost interactions without restrictive parametric assumptions. The informative censoring problem is handled using inverse probability of censoring weighted estimating equations. The proposed method is illustrated using simulation and also with data on the cost of dialysis for patients with end-stage renal disease.

KEYWORDS: Cost-benefit analysis; Economic evaluation; Informative censoring; Quality of life; Survival analysis.

*AK Jain is doctoral student, University of Michigan, Ann Arbor, MI 48109-2029. RL Strawderman is Associate Professor, Department of Biometrics, Cornell University, Ithaca NY 14853.

1 Introduction

Cost measurement can be roughly defined as measuring resource consumption, in terms of dollars, for a given medical intervention. Cost measurement in health care has become increasingly popular in recent decades and shows no signs of abating; a critical look at current trends is available in Triplett (1999). This trend of measuring costs results from society's willingness to pay for health care interventions. Given budgetary constraints, we must wisely allocate resources in order to maximize health benefits for individuals. Gaining insight into patient-level factors that affect the cost of medical interventions is therefore crucial if we are to devise cost-effective strategies for prevention and intervention.

This paper is concerned with the problem of modeling lifetime (i.e., cumulative) medical costs measured from some well-defined point of patient entry (e.g., start of treatment) to some well-defined event (e.g., death). In follow-up studies with finite time horizons, it is inevitable that some individuals under study are censored prior to observing the event of interest. In such cases, lifetime costs are also censored, and an important issue emerging in the biostatistics literature is the proper handling of such data. In particular, even in cases where censoring on the time scale is purely administrative (i.e., random), censored lifetime cost endpoints are subject to "induced" informative censoring. Lin *et al.* (1997) describe this problem in the context of patient-level cost functions. Individuals who accrue costs at higher (lower) rates will tend to have larger (smaller) cumulative costs at both censoring and event times. Consequently, the transformation of time by the individual cost functions induces dependence between the respective values of the cost process at the censoring and event times. This dependence exists even if the censoring and event times are themselves independent.

Similar problems arise in other settings involving endpoints that may be viewed as random transformations of survival times. For example, Gelber *et al.* (1989) discuss the dependence between the censored and uncensored quality-adjusted lifetime measure TWiST. Zhao and Tsiatis (1997) formulate and solve a related problem that deals with the nonparametric analysis of quality-adjusted survival time data. In fact, problems of induced informative censoring occur naturally in cases where the process of interest (e.g. cumulative cost) is increasing over time and its observation is stopped due to the occurrence of a possibly dependent terminal event (Strawderman, 2000). In such cases traditional survival analysis tools (e.g., Kaplan-Meier, Cox proportional hazards models, etc ...) that assert independence between censoring and the outcome variable are no longer appropriate.

A handful of statistical methods appropriate for dealing with lifetime cost data, and more generally response variables subject to induced informative censoring, have been proposed. Essentially all published work has dealt with the problem of nonparametric estimation, ignoring important covariate information; see, for example, Lin *et al.* (1997), Zhao and Tsiatis (1997), Cook and Lawless (1997), van der Laan and Hubbard (1999), Bang and Tsiatis (2000b), and Strawderman (2000). Considerably less work has been

done leading to valid estimators that adjust for patient covariate information. Bang and Tsiatis (2000a) consider median regression for lifetime cost data assuming censoring is independent of all else, including covariates. Lin (2000a) considers proportional-means regression models under similar assumptions. Lin (2000b) considers direct extensions of the linear regression model, and allows censoring to further depend on covariates. The fundamental difference between Lin (2000a) and Lin (2000b) is whether one prefers to allow for multiplicative versus additive covariate effects on lifetime costs. All of the methods proposed thus far entail restrictive assumptions regarding the relationship between lifetime cost and covariate information.

In this paper we focus on modeling the lifetime cost distribution as a function of patient covariates. Kooperberg, Stone, and Truong (1995a) develop and implement hazard regression (HARE) modeling for noninformatively censored data by modeling the log hazard function using linear splines and their tensor products. We adapt their approach to informatively censored lifetime cost data by employing appropriate "inverse probability of censoring weighted" (IPCW) estimating equations derived from those of the original HARE model. This modeling paradigm, referred to as IPCW-HARE, proves versatile in its lack of restrictive assumptions (e.g., proportional hazards, a linear relationship between survival time and covariates, etc ...) and in its user-friendly software implementation. Further development and discussion of issues unique to IPCW-HARE are given in Section 2. A fundamental difference between IPCW-HARE and the methods of Lin (2000a,b) and Bang and Tsiatis (2000a) is that we have elected to model the conditional distribution function of lifetime cost given covariates rather than a single summary measure (e.g., mean or median cost). This approach affords certain advantages. For example, in addition to being able to compute various summary measures (e.g., mean or median costs), one may gain further insight by exploring the conditional hazard, density, and cumulative distribution function of costs for different covariate patterns. In Section 3, we evaluate the utility of IPCW-HARE for estimating mean and median costs using simulation and compare our results to the estimated mean and median costs respectively obtained using the methods of Lin (2000b) and Bang and Tsiatis (2000a). Finally, in Section 4 we analyze some cost data associated with two common modes of dialysis for Medicare patients with end-stage renal disease.

Although we focus here on modeling lifetime medical cost data, the present framework extends without alteration to other settings involving continuous outcomes. For example, our methods can be used for quality of life studies, where censoring induces a dependence between the censored outcome process and the actual outcome (e.g., Zhao and Tsiatis, 1997). However, because the focus is on modeling the hazard function of an absolutely continuous random variable, the methods to be discussed here are not appropriate for discrete outcomes. Thus, for example, the present modeling framework would not be appropriate in the recurrent event setting considered by Cook and Lawless (1997) since the lifetime cumulative number of events is a discrete outcome variable.

2 HARE and its IPCW extension

2.1 Notation and Assumptions

In order to discuss the notion of "lifetime cost", we must first define a notion of "lifetime". Similarly to Lin (2000a,b), we assume interest lies in costs accrued over the period $[0, L]$, where $L < \infty$. Patients may experience a terminal event U (e.g., death) prior to L, and we shall subsequently interpret "lifetime" as being the minimum of these two times. More precisely, let $T = \min(U, L)$ denote "lifetime", and define the "lifetime cost" V to be the cost accumulated over $[0, T]$. Evidently, V is the terminal cost if $U \leq L$ and is (say) the $L-$year cost otherwise. The role of L, as well some restrictions on its selection, are discussed further in the last paragraph of this section.

As described in the Introduction, the event T, hence V, will not be observed for all patients. For a given patient, let C and D respectively denote a potential censoring time and corresponding censored value of lifetime cost. We suppose that (i) if $T \leq C$, then $V \leq D$; and, (ii) if $T > C$, then $V > D$. A standard convention in survival analysis is that only $X = \min(T, C)$ and $\Delta = I\{T \leq C\}$ may be observed for a given patient. Consequently, observable data on any given patient are assumed to take the form (Y, X, Δ, Z), where $Y = \min(V, D)$ and Z represents a vector of baseline covariates. Note in particular that Δ serves as the censoring indicator for both lifetime and lifetime cost. Our observed data are assumed consist of a random sample of the form $(Y_i, X_i, \Delta_i, Z_i)$, $i = 1 \ldots n$.

We have chosen to model the hazard function of V given Z; hence, it is explicitly assumed that V is a continuous random variable. For simplicity, it is further assumed that (1) T and C are also continuous random variables; (2) the covariates Z are bounded; (3) T and C are independent given Z; and, (4) V and C are independent given Z. Importantly, however, no assumptions are made regarding the dependence between T and V or T and Z. Technical considerations further dictate that L must satisfy $K(L|Z) > 0$ with probability one, where $K(c|z) = P(C > c|Z = z)$. To better understand this condition, suppose L were equal to the largest possible terminal event time U (i.e., $L = \sup\{t : P\{U > t\} > 0\}$) and that $K(L|z) = 0$ for all possible realizations z of Z. This implies all patients are censored prior to L; consequently, without further assumptions on the rate at which costs accrue over time for different patients, it would be impossible to estimate the total cost on $[0, L]$.

2.2 HARE: A Brief Review

Kooperberg *et al.* (1995a) propose a hazard regression (HARE) model for positive, right-censored time-to-event data. In particular, they assume

$$\log \lambda_\beta(v|z) = \sum_{j=1}^{p} \beta_j B_j(v|z) I\{v > 0\}, \tag{1}$$

where $\lambda_\beta(v|z)$ denotes the conditional hazard function for the uncensored outcome variable (e.g., V), $B_1 \ldots B_p$ form a basis for a space of functions defined in V and Z, and β_j is the regression coefficient associated with the j^{th} basis function. Essentially, the functions $B_j, j = 1 \ldots p$ together define linear or spline functions of the outcome and/or continuous covariates, linear functions of categorical covariates, and two way interactions. A full description of the basis, as well as the space of functions spanned by (1), can be found in Kooperberg *et al.* (1995a, §5). The HARE model (1) is exceptionally flexible in its use of piecewise linear splines and does not *apriori* make restrictive assumptions about the relationship between V and Z (e.g., proportional hazards). Notably, the HARE model does contain the (parametric) proportional hazards model as a special case since, under (1), $\lambda_\beta(v|z) = \lambda_0(v)g(z)$ (say) if none of the basis functions $B_j(v|z)$ simultaneously depend on v and z.

Kooperberg *et al.* (1995a) fit (1) to randomly right-censored data using maximum likelihood, the relevant partial likelihood being

$$\prod_{i=1}^{n} \lambda_\beta(Y_i|Z_i)^{\Delta_i} \bar{F}_\beta(Y_i|Z_i), \tag{2}$$

where $\bar{F}_\beta(y|z)$ is the survivor function associated with the hazard function $\lambda_\beta(y|z)$. Employing a notion of "allowable spaces", they utilize a stepwise addition and deletion procedure for choosing basis functions that represents a hybrid of well-known stepwise addition and deletion procedures appropriate for linear and generalized linear models. The Bayesian Information Criterion (BIC; see Schwarz, 1978) is used to select the final model from the sequence of fitted models. We refer the reader to Kooperberg *et al.* (1995a) for further details regarding the model fitting and selection procedures, and Kooperberg, Stone, and Truong (1995b) for some relevant asymptotic theory.

2.3 IPCW-HARE: HARE for informatively censored data

The likelihood (2) is valid assuming censoring is noninformative. As discussed earlier, this assumption is violated when analyzing censored lifetime cost data. In particular, if one applies HARE directly to the observed data $(Y_i, X_i, \Delta_i, Z_i)$, $i = 1 \ldots n$, an inconsistent estimate of the hazard function, hence desired summaries (e.g., means, medians, etc ...), is the likely result. In this section, we propose a modification of HARE in order to account for the informatively censored nature of lifetime medical cost data. Importantly, the model (1) is not being changed; rather, the method by which (1) is estimated is what requires modification.

To motivate the basic IPCW-HARE estimating function, suppose there is no censoring. Then, the contribution of the i^{th} individual to the loglikelihood function for $\beta = (\beta_1 \ldots \beta_p)^T$ takes the following exponential family form (Kooperberg *et al*, 1995a):

$$\ell_i(\beta) = \left(\sum_{j=1}^{p} \beta_j B_j(Y_i|Z_i) \right) - \int_0^{Y_i} \exp \left(\sum_{j=1}^{p} \beta_j B_j(u|Z_i) \right) du \tag{3}$$

for $Y_i \geq 0$; note that $Y_i = V_i$ here since we have assumed there is no censoring. The resulting score vector contribution is $S_i(\beta) = (S_{i1}(\beta) \ldots S_{ip}(\beta))^T$, where

$$S_{ij}(\beta) = B_j(Y_i|Z_i) - \int_0^{Y_i} B_j(u|Z_i) \exp\left(\sum_{j=1}^{p} \beta_j B_j(u|Z_i)\right) du. \tag{4}$$

Based on a random sample of size n, the MLE of β is obtained via the score equation $S_{full}(\beta) = \sum_{i=1}^{n} S_i(\beta)$. In the missing data literature, $S_{full}(\beta)$ is generally referred to as the "full data" score function, an appropriate characterization since it is derived from the full data (V, T, Z). Obviously, the MLE of β cannot be computed via $S_{full}(\beta)$ in the presence of censored data.

In a rather general missing data setting, Robins and Rotnitzky (1992) propose to construct estimating functions from complete observations by inversely weighting "full data" contributions by the respective probability of being observed. In the context of the present problem, and assuming that our observed data take the form $(Y_i, X_i, \Delta_i, Z_i)$ $i = 1 \ldots n$, a valid IPCW estimating equation for β is

$$S^*_{ipcw}(\beta) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{K(X_i|Z_i)} S_i(\beta), \tag{5}$$

the normalization of n^{-1} being used for notational convenience only. By reweighting score contributions using this IPCW scheme, we are calculating an M-estimator for β that accounts for the informatively censored nature of the data. In fact, under mild conditions, (5) is easily shown to be an unbiased estimating function for β; see, for example, Robins and Rotnitzky (1992).

Since $K(\cdot|\cdot)$ is generally unknown, an estimate of $K(\cdot|\cdot)$ must be substituted into (5) in practice. In the case where the censoring process is not influenced by covariates (e.g., administrative censoring) this is easy to do using the Kaplan-Meier estimator. If censoring does depend on covariates then a model must be chosen. A simple and attractive choice here is to assume that the censoring mechanism follows a Cox proportional hazards model. In addition to being widely used and well understood, this model contains the important case of independent censoring. Whatever estimate of $K(\cdot|\cdot)$ the user employs, the resulting estimating function takes the form

$$S_{ipcw}(\beta) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(X_i|Z_i)} S_i(\beta), \tag{6}$$

and the estimate $\widehat{\beta}$ is that which solves $S_{ipcw}(\beta) = 0$. Since (5) and (6) coincide when $\widehat{K}(\cdot|\cdot) = K(\cdot|\cdot)$, we only consider (6) from this point forward.

The estimating function (6) forms the core of IPCW-HARE, and replaces the score function derived from (2) that is used by HARE. We now briefly outline some relevant asymptotic properties of $\widehat{\beta}$. Precise statements of regularity conditions are purposely avoided since the asymptotic results provided below do not account for the data-driven fitting procedures employed by IPCW-HARE and thus are not useful for statistical inference. Rather, these results are used primarily in §2.4, where some important

changes to the model building and model selection criteria originally employed by HARE are discussed in detail. These changes represent a necessary consequence of the move away from a maximum likelihood estimation framework.

Suppose the set of basis functions $B_j(\cdot|\cdot)$, $j = 1 \ldots p$, hence the dimension of the parameter space, is fixed and finite. Let $\beta^* \in I\!R^p$ be the unique maximizer of the strictly concave function $E[\ell_1(\beta)]$ with $\ell_1(\beta)$ being given by (3). In the appendix, consistency of $\hat{\beta}$ for β^* is established provided the estimated weights satisfy certain consistency conditions. Let $S'_{ipcw}(\beta)$ denote the derivative of $S_{ipcw}(\beta)$ with respect to β, and suppose that $-S'_{ipcw}(\beta^*) \xrightarrow{P} M(\beta^*)$, where $M = M(\beta^*)$ is nonsingular. Then, under suitable conditions, it follows that

$$\sqrt{n}\left(\widehat{\beta} - \beta^*\right) \xrightarrow{d} N(0, M^{-1}QM^{-1})$$

where Q is the asymptotic variance of (6) at β^*. This result holds whether or not $K(\cdot|\cdot)$ is estimated from the data, with the effect of estimating $K(\cdot|\cdot)$ being entirely reflected through Q. Specifically, if $K(\cdot|\cdot)$ is known,

$$Q = E\left[\frac{\Delta}{K^2(T|Z)} S_1(\beta) S_1^T(\beta)\right];$$

in contrast, when $K(\cdot|\cdot)$ is estimated from the data, the actual form of Q depends on the model and method used for estimating $K(\cdot|\cdot)$. In the appendix, a specific formula for Q is devised for the case where $K(\cdot|\cdot)$ follows a (semiparametric) Cox proportional hazards model and is estimated accordingly.

The criteria used by IPCW-HARE in building models are discussed in detail in the next section. These criteria depend in part on the assumed asymptotic behavior of $\widehat{\beta}$, and in particular on the matrices Q and M. The contrast in asymptotic behavior that results from knowing versus estimating $K(\cdot|\cdot)$ has some important practical implications. Specifically, suppose first $\widehat{K}(\cdot|\cdot)$ is regarded as an estimated quantity. Then, the dependence of Q on the model for the censoring mechanism implies that different censoring models (e.g., a semiparametric versus fully parametric Cox proportional hazards model) necessitate completely separate implementations of IPCW-HARE. Such dependence further rules out the use of HARE and other data-adaptive methods for estimating $K(\cdot|\cdot)$ since their asymptotic behavior cannot be characterized precisely. The situation changes dramatically when $\widehat{K}(\cdot|\cdot)$ is regarded as a known quantity (i.e., treated as known, even if it is estimated). In particular, since Q no longer depends on the assumed model for $K(\cdot|\cdot)$, it may be estimated using $\widehat{Q} = n^{-1}\sum_i \widehat{\psi}_i \widehat{\psi}_i^T$, where $\widehat{\psi}_i = \frac{\Delta_i}{\widehat{K}(X_i|Z_i)} S_i(\widehat{\beta})$. Because computation of $\widehat{Q}$ only requires the estimated weights (i.e., as opposed to knowledge of the model that generated them), the same implementation of IPCW-HARE may be used however $K(\cdot|\cdot)$ is estimated.

For a fixed set of basis functions, the implications of treating the estimated quantity $\widehat{K}(\cdot|\cdot)$ as known are clear: while the consistency of $\widehat{\beta}$ is unaffected, fixed-size tests and fixed-level confidence sets will respectively have incorrect size and coverage. The decision rules employed by IPCW-HARE during the model building process do not rely on formal tests of significance or fixed level confidence sets, and

$\widehat{K}(\cdot|\cdot)$ is in fact held fixed throughout the entire model-building phase. Consequently, in the context of adaptively building a model for a given dataset, the need to acknowledge the fact that $\widehat{K}(\cdot|\cdot)$ is estimated is less clear. Moreover, the possible limitations that result from ignoring the fact that $\widehat{K}(\cdot|\cdot)$ is estimated are offset by very substantial practical gains. For example, in addition to allowing the user to employ any reasonable method for estimating $K(\cdot|\cdot)$, tremendous gains in computational efficiency are realized. Since standard errors will be computed via nonparametric bootstrap (i.e., resampling $(Y_i, X_i, \Delta_i, Z_i)$, $i = 1 \ldots n$ with replacement), such gains in computational efficiency become important, particularly for large datasets. Bootstrapping is also advantageous since the standard deviation estimates reflect both estimation of the weights and the adaptive model selection scheme.

2.4 IPCW-HARE: Important modifications to HARE

Despite our reweighting of the HARE score equations, many of the procedures described in Kooperberg *et al.* (1995a) used for fitting (1) to data (e.g., knot placement, starting values, etc ...) were able to be utilized by IPCW-HARE without significant alteration. However, some important modifications to HARE were also required. Essentially, these modifications were needed in two places: adding and deleting basis functions and selecting the "best" HARE model. For reasons discussed earlier, our particular implementation of IPCW-HARE treats $\widehat{K}(\cdot|\cdot)$ as known for the purposes model building; consequently, in what follows we employ $\widehat{M} = -S'_{ipcw}(\widehat{\beta})$ and $\widehat{Q} = n^{-1}\sum_i \widehat{\psi}_i\widehat{\psi}_i^T$, where $\widehat{\psi}_i = \frac{\Delta_i}{\widehat{K}(X_i|Z_i)}S_i(\widehat{\beta})$. The changes to be described below drastically reduced the incidence of overfitting (i.e., the propensity of HARE to fit the model to extreme data points), and as our simulation results will show, yield estimators with good overall performance.

2.4.1 Stepwise Addition and Deletion of Basis Functions

During the addition phase of the model building process, HARE attempts to enrich the class of hazard functions covered by (1) by moving from (say) a $p-1$ dimensional "allowable space" G_0 to a p dimensional allowable space G that contains G_0. In practical terms, this is handled by adding a candidate basis function B_p, hence coefficient β_p, to the model. At any given step, there are a potentially large number of candidates; consequently, due to the computationally intensive nature of the fitting process, HARE cannot refit the model and compare loglikelihoods like one might do with a generalized linear model. Instead, HARE "tests" the hypothesis that (1) is a member of G_0 using a Rao statistic, and in particular adds the basis function which maximizes this statistic (or minimizes the p-value). Kooperberg *et al.* (1995a, §4) provide further details.

Like HARE, IPCW-HARE starts with a constant hazard function and adaptively builds a sequence of larger models. The version of the Rao statistic used by HARE is not really appropriate for use by

IPCW-HARE because it uses the inverse information matrix as an estimate of variance. Let $\widetilde{\beta}_0$ denote the solution obtained by setting (6) equal to zero for the model corresponding to G_0. Suppose the test we are conducting is for $\beta_p = 0$. Let $S_G(\beta)$ denote (6) for a candidate hazard model corresponding to space G, let $-\widehat{M}_G(\beta)$ be its first derivative, and let $\widehat{Q}_G(\beta)$ denote an estimate of $\mathrm{Var}[S_G(\beta)]$. Finally, let $S_{Gp}(\beta)$ denote the p^{th} element of $S_G(\beta)$. Then, the Rao statistic we use in deciding whether or not to add a basis function is

$$R^2 = nS_{Gp}(\widehat{\beta}_0)^T \widehat{C}^{-1} S_{Gp}(\widehat{\beta}_0)^T,$$

where $\widehat{\beta}_0 = (\widetilde{\beta}_0, 0)^T$ and $\widehat{C}$ is computed via (5) of Heritier and Ronchetti (1994) with (in their notation) $\mathbf{M} = \widehat{M}_G(\widehat{\beta}_0)$ and $\mathbf{V} = \widehat{M}_G^{-1}(\widehat{\beta}_0)\widehat{Q}_G(\widehat{\beta}_0)\widehat{M}_G^{-1}(\widehat{\beta}_0)$. The basis function added at a given step is then that with the largest value of R^2. The Rao statistic R^2 reduces to that used by HARE if $\mathbf{V} = \mathbf{M}^{-1}$, as would be the case under maximum likelihood estimation.

Kooperberg *et al.* (1995a, §11.4) list 3 criteria used by HARE for deciding whether to continue adding basis functions. Two of these have clear practical justifications; the last essentially corresponds to a likelihood-ratio-based criterion for assessing improvement in model fit. Specifically, suppose the present model has P basis functions. Then, HARE stops adding basis functions if $2(\widehat{\ell}_P - \widehat{\ell}_p) < (P - p) - 1$ for some p with $2 \leq p \leq P - 3$, where $\widehat{\ell}_P$ and $\widehat{\ell}_p$ respectively represent the (estimated) loglikelihood for a model with P and p basis functions. Since the fitting process ensures that the model with p basis functions is nested within the model with P basis functions, $2(\widehat{\ell}_P - \widehat{\ell}_p)$ is approximately distributed as χ^2_{P-p} (i.e., assuming that the model with p basis functions is sufficient). Thus, HARE discontinues the addition phase if the change in loglikelihood values does not exceed its mean value (minus one).

REMARK: Asymptotically, HAREs criteria boils down to unnecessarily adding a basis function with probability $P\{\chi^2_{P-p} > P - p - 1\}$. Since $P - p \geq 3$, it is easy to show that this probability is at most 0.57, and monotonically decreases to 1/2 as $P - p$ get large. Hence, the decision to add a basis function when it is in fact not needed is roughly equivalent to tossing a (slightly) biased coin.

Since IPCW-HARE is not maximum likelihood, HAREs particular criteria for assessing model fit carries questionable relevance. Let $\ell_i(\widehat{\beta}^{(p)})$ and $\ell_i(\widehat{\beta}^{(P)})$ denote the contribution of patient i to the *full data* loglikelihood (3) for nested models respectively having p and P basis functions; note that these are to be interpreted in a similar manner to $\widehat{\ell}_p$ and $\widehat{\ell}_P$. Consider the likelihood-ratio-type statistic

$$L_n^2 = 2\sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(X_i|Z_i)} \left(\ell_i(\widehat{\beta}^{(P)}) - \ell_i(\widehat{\beta}^{(p)})\right);$$

see, for example, Heritier and Ronchetti (1994, eqn. 6). If $K(\cdot|\cdot)$ were known, L_n^2 could be derived as the likelihood ratio statistic assuming the data follow a density proportional to $\exp(-\tau(Y, X, \Delta, Z, \beta))$, where $\tau(Y, X, \Delta, Z, \beta) = \Delta\ell(\beta)/K(X|Z)$ and $\ell(\beta)$ is computed similarly to (3) (cf. Ronchetti, 1997, §3.1). Consequently L_n^2 represents a reasonable statistic for assessing improvement in model fit. However, the (asymptotic) mean of this statistic is not $P - p$. In fact, under suitable conditions, L_n^2 behaves

asymptotically like a finite mixture of χ^2_1 random variables, the mixing coefficients being governed by the eigenvalues of a certain matrix depending on Q and M defined in Section 2.3 above; see Heritier and Ronchetti (1994, §2 and Proposition 3a) for further details. Analogously to HARE, the criteria we use for stopping the addition of basis functions is $L_n^2 < \widehat{\mu}_{P-p} - 1$, where $\widehat{\mu}_{P-p}$ denotes an estimate of the mean of the appropriate mixture of chi-square random variables.

Finally, upon stopping stepwise addition using the rule described above, we begin stepwise deletion. The decision rule for keeping the p^{th} basis function is based on the Wald statistic

$$W^2 = \frac{\widehat{\beta}_p^2}{\widehat{V}_{pp}}$$

where $\widehat{V} = n^{-1}\widehat{M}_G^{-1}(\widehat{\beta})\widehat{Q}_G(\widehat{\beta})\widehat{M}_G^{-1}(\widehat{\beta})$ and $\widehat{\beta}$ is the solution to (6) under model G. In particular, for a given model dimension, the basis function corresponding to the smallest value of W^2 is deleted. Similarly to the stepwise addition phase, this criteria reduces to that used by HARE if $M_G = Q_G$, and the process of deleting basis functions is continued until the smallest possible model is reached.

2.4.2 Model Selection

The stepwise process of adding and deleting basis functions creates a sequence of models for the hazard function, one of which must be selected. HARE employs the Bayesian Information Criterion (BIC; see Schwartz, 1978) for this, and chooses between models using penalized comparisons between loglikelihoods. For reasons similar to those described in the last section, a modification is employed due to the absence of an appropriate likelihood. Ronchetti (1997) develops a Robust Akaike Information Criterion (AICR) for the M-estimator setting. Let $\widehat{\beta}$ denote the parameter estimate obtained under a given model, and consider (cf. Ronchetti, 1997, eqn. 3.2)

$$AICR = 2\sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(X_i|Z_i)} \cdot \ell_i(\widehat{\beta}) - \alpha \log(n_u)\mathrm{tr}(\widehat{M}^{-1}\widehat{Q}),$$

where $\widehat{M}$ and $\widehat{Q}$ denote estimates of M and Q evaluated at $\widehat{\beta}$, n_u denotes the number of uncensored observations, and α is a fixed multiplier. Roughly speaking, the role of AICR is similar to BIC, and model selection by IPCW-HARE is carried out by selecting the model with the maximum value of AICR, the default choice of α being 1. The use of $log(n_u)$ as the penalty parameter is meant to emphasize the fact that the amount of information present in censored data substantially depends on the number of uncensored observations rather than total sample size.

3 Simulation Results

To evaluate the performance of IPCW-HARE we conducted Monte Carlo simulations with various lifetime cost schemes. IPCW-HARE approximates the hazard function of the cost variable as a function of

covariates; for simplicity we report only the mean and median lifetime cost estimates derived from this hazard estimate for specific covariate patterns. We fit restricted models, in which no covariate-by-cost interactions are allowed (i.e., proportional hazards, or PH=T); we also fit unrestricted models (PH=F) in which covariate-by-cost interactions are allowed. Importantly, in none of our examples is there an obvious "true" model, meaning that the resulting data generation scheme is in exact correspondence with a member of the parametric family of models we are working within. Hence, in each case the hazard (1) at best represents an approximation to reality, and these simulation results constitute an illustration of the "real world" performance of the IPCW-HARE procedure.

Since a nice analytic expression for the mean cost under the IPCW-HARE model is not available, we estimate this using Monte Carlo. In particular, for a given hazard function estimate, we generate 2,000 random lifetime cost observations using HAREs built-in function rhare, and then take a simple average of these to obtain our Monte Carlo estimate of the mean lifetime cost. To estimate the median cost, we again use the built-in functionality of HARE by employing the qhare function provided with the original HARE software. The censoring weights are computed using the product-integral form of the survivor function appropriate for a Cox regression model; see Andersen *et al.* (1993, eqn. 7.2.34) for details. All simulations employ 2,000 replicated datasets of size $n = 200$.

Our first simulation is adopted from Bang and Tsiatis (2000a). Survival and censoring times (in years) respectively follow Uniform[0,10] and Uniform[0,12.5] distributions. We set $L = 10$ years, which leads to approximately 40% censoring. A single covariate Z is assumed Uniform[20,70]. The cost function for an individual i consists of a one-time baseline diagnostic cost B_{Oi}, an annual cost B_{ij} uniformly distributed throughout each j^{th} year, and a death cost D_i uniformly distributed over the last year of life. Specifically, $B_{Oi} = 500 + 100Z_i + \varepsilon_i$, $B_{ij} = 400 + 10Z_i + \varepsilon'_{ij}$, and $D_i = 1000 + 200Z_i + \varepsilon''_i$ where ε_i is Uniform[2500,7500], ε'_{ij} is Uniform[500,1300], and ε''_i is Uniform[5000,15000]. Notice that the covariate only affects cost, and does so linearly; survival and censoring variables do not depend on this information. Although censoring is independent of the covariates, the weights are still estimated using a Cox proportional hazards model with Z as the lone covariate. In order to assess the effect of including weights (and hence of informative censoring on cost), we have included results obtained from fitting the original (i.e., unweighted) HARE model to these data for comparison. Finally, we have also included the estimators of Lin (2000b) and Bang and Tsiatis (2000a), which ought to perform very well here due to the structure of the cost function.

The results of the first simulation are summarized in Tables 1 and 2. IPCW-HARE is seen to perform well in estimating both mean and median lifetime cost across the range of Z. In contrast, HARE demonstrates a substantial negative bias for all covariate patterns, the largest bias being on the order of \$1,900. Consequently, there is a significant (and predictable) effect of informative censoring in this example. The fact that IPCW-HARE leads to more variable estimates is expected; HARE uses both

censored and uncensored cost observations, while IPCW-HARE only uses uncensored cost observations. Consequently, HARE, though biased, makes more efficient use of the available data. The estimators of Lin (2000b) and Bang and Tsiatis (2000b) generally outperform those produced by IPCW-HARE for respectively estimating the mean and median costs. This is also to be expected since these methods are specifically designed for estimating the effects of covariates with a linear relationship to lifetime cost. Finally, it is seen that the allowance for interactions between cost and covariates makes little difference in the IPCW-HARE estimates, indicating a proportional hazards assumption is reasonable here.

The first simulation assumed costs accrue uniformly on a yearly basis, and that costs increase linearly as a function of a single covariate. In our second simulation, we instead allow the cost-incurring episodes to occur at random times; that is, we assume recurrent events with associated costs. The first covariate Z_{1i} is Uniform[10,50] and the second covariate Z_{2i} is Bernoulli with $p = \frac{1}{3}$. Given $Z_{1i} = z_1$, survival times are Gamma distributed with a conditional mean of $\frac{12}{30}z_1$ years; the unconditional mean survival time is 12 years. Censoring times are Exponential with a mean of 18 years and L=15. Conditional on failure time and covariate information, the recurrent events follow a homogeneous Poisson process with rate $\lambda(s|T_i = t, Z_{1i} = z_1) = 0.01 \cdot z_1 \cdot \log(\max(3, t))$ and the cost of each event is distributed as Uniform[\$2000,\$5000]. Initial costs for patients, or B_{Oi}, are assumed Uniform[\$0,\$5000] if $Z_{2i} = 0$ and Uniform[\$5000,\$10000] if $Z_{2i} = 1$. Finally, given $Z_{1i} = z_1$, the death cost D_i is \$10000 Uniform[1,$f(z_1)$] where $f(z_1) = \sqrt{z_1/10} \in (1, 2.24)$, and is uniformly distributed over the last two years of life. This leads to an average of 7.6 cost-incurring episodes per patient over the course of their (truncated) lifetime and 34% censoring. Importantly, the covariate Z_2 affects baseline costs only, and does so linearly; the covariate Z_1 affects both survival and cost in a nonlinear way. Censoring is independent of the covariate; however, the weights are estimated using a Cox proportional hazards model based on both Z_1 and Z_2.

The results of the second simulation are given in Tables 3 and 4. The mean and median of the covariate Z_1 both occur at 30. Comparing the IPCW-HARE models "PH=F" versus "PH=T", we see the former tends to exhibit substantially less bias, indicating proportional hazards may not be a reasonable assumption. Across the range of Z, IPCW-HARE (PH=F) is seen to outperform the simple regression-based estimators for estimating both mean and median costs. Although the Bang and Tsiatis (2000a) estimator does very well for $Z_1 = 15$, its performance degrades significantly as Z_1, hence the level of censoring, increases. Comparing Table 4 and Table 2, we further see that the efficiency gains of these regression methods over IPCW-HARE are significantly reduced when the true cost function and linear regression model fail to coincide.

In our third and final simulation, we move away from the previous two paradigms and simulate lifetime cost directly. That is, we employ a model to generate only the lifetime cost V_i, rather than computing it indirectly via accumulating intermediate costs. We also allow for both survival and censoring to depend on relevant covariates. The covariate Z_{1i} is Uniform[10,70] and Z_{2i} is Bernoulli with $p = \frac{1}{2}$. If

$10 \leq Z_{i1} < 30$, the survival time U_i is Exponential with mean 20 years; if $30 \leq Z_{i1} < 50$, U_i is Exponential with mean 15 years; and if $50 \leq Z_{i1} \leq 70$, U_i is Exponential with mean 10 years. The censoring time C_i is Exponential with mean $25Z_{2i} + 18(1 - Z_{2i})$ years. We set L=25 and $T_i = \min(U_i, L)$. This leads to approximately 39% censoring overall. Finally, the terminal cost V_i is assumed to follow a Gamma distribution with mean $\alpha(Z_i)\beta(T_i, Z_i)$ and standard deviation $\sqrt{\alpha(Z_i)}\beta(T_i, Z_i)$. The shape parameter $\alpha(Z_i) = \frac{1}{3}Z_{2i}Z_{1i}^2 + \frac{2}{3}(1-Z_{2i})Z_{1i}^2$ and the scale parameter $\beta(T_i, Z_i) = Z_{2i}\log(1+T_i) + 2(1-Z_{2i})\log(1+T_i)$. This implies, given T_i and Z_i, that the average cost equals $\frac{1}{3}(2-Z_{2i})^2 Z_{1i}^2 \log(1+T_i)$ and that the standard deviation equals $\frac{1}{\sqrt{3}}(2 - Z_{2i})^{3/2} Z_{1i} \log(1 + T_i)$. Consequently, mean lifetime costs generally rise (and become more variable) as Z_{1i} rises; the rates at which these costs rise (and their level of variability) depends on Z_{2i}. Consequently, there is a rather complicated interaction between the effects of Z_1, Z_2 and T on the terminal cost V.

The results of the third simulation are given in Tables 5 and 6. In terms of bias, IPCW-HARE performs very well in estimating both mean and median lifetime cost across the various covariate combinations. In contrast, the linear regression models perform rather poorly, in some cases having enormous bias. In fact, for the covariate combination (15,1), these regression models led to negative simulated average mean and median lifetime costs. To be fair, these linear regression methods might perform better if, for example, costs were placed on the log scale, interaction terms were placed in the model, and nonlinear covariate effects were allowed. This was not done to better illustrate the point that IPCW-HARE is able to produce reasonable estimates with much less preliminary analysis.

This suite of simulations suggests that IPCW-HARE performs well under a variety of situations, and show in particular that (i) IPCW-HARE-based estimates tend to be less biased than those obtained using methods that impose stronger assumptions; and, (ii) this increase in accuracy usually comes at some expense in precision. Our results also indicate that the ignoring the censored cost observations entirely probably entails a significant loss in precision (i.e., see Table 2); we return to this issue in the Discussion. We suspect that some portion of the residual biases observed in Tables 1, 3, and 5 are a result of model misspecification. Specifically, despite being adaptive, IPCW-HARE eventually yields estimates able to accurately approximate functions that lie within some restricted class only. Consequently, in finite samples model misspecification is still an important consideration. Reducing such bias should be possible by increasing the flexibility of the basis; for example, by employing using cubic instead of linear splines.

4 End-Stage Renal Disease Data

The data consist of subjects with End-Stage Renal Disease (ESRD) collected from the United States Renal Data System (USRDS). Individuals were included in the study if they started peritoneal dialysis

or hemodialysis within the window of January 1, 1992 thru December 31, 1996. All subjects are 66 and older, have Medicare as a primary payer, and are possibly censored due to loss-to-follow-up, end of study (12/31/96), or kidney transplant. Available covariates include age at the start of treatment (66-70, 71-75, and > 75), gender, whether diabetes is the primary cause of ESRD, and race (white, black, and other). There are 38,732 subjects available for analysis; however, because IPCW-HARE is computationally intensive, we have elected to use four random sub-samples ($n = 1,000$ from each dialysis $\times$ diabetes status category) for this illustration. The reason for this really rests more in the need to compute standard errors via bootstrap than it does in fitting the model itself. A comparison of summary statistics between these sub-samples and the entire data set indicates that our subsamples are representative samples.

For this analysis we focus on estimating mean lifetime dialysis costs for various patient profiles; for example, 66-70 year old diabetic white males on hemodialysis vs. 66-70 year old diabetic white males on peritoneal dialysis. We have truncated the time scale to 4 years (i.e., $L = 4$ years) in order to ensure that our "probability of censoring" weights remain positive. Hence we are really estimating the four-year costs for different modes of dialysis. Since nearly 80% of dialysis patients survive less than four years (USRDS, 1999), we still refer to this estimate as the mean lifetime cost.

The data were stratified by dialysis modality and diabetic status and then analyzed separately. Within each strata, our regression model was fit using the following dichotomous covariates: gender, middle-old (71-75), old-old (> 75), black, with the respective baseline categories taken to be male, 66-70, and white. For this analysis, the race category "other" is excluded; in each case these patients comprise less than 5% of the sample, and even less in terms of the number of deaths. Since all covariates are dichotomous, splines are being fit to costs only, with interactions between pairs of covariates and with cost being selected adaptively. The standard deviation of the resulting point estimates were approximated using 1000 bootstrap samples.

Table 7 summarizes the results of this analysis, and certain interesting patterns emerge. For example, at all reported age/race combinations, blacks exhibit higher average treatment costs than whites; also, controlling for diabetic status, hemodialysis is more costly than peritoneal dialysis. Generally, lifetime treatment costs do not appear to differ much by gender, and no clear trend emerges for the costs of treating diabetic versus non-diabetic patients. The estimated bias of these point estimates, obtained via bootstrap approximation, ranges from -\$10,000 to \$14,000 for hemodialysis point estimates and from -\$4,400 to \$11,700 for peritoneal dialysis point estimates. Except for black hemodialysis patients, histograms of these mean estimates were observed to be unimodal and approximately normally distributed. The lack of multimodality indicates that estimated covariate effects largely remained stable across the bootstrap iterations. The mean estimates for black hemodialysis patients were bimodal, the second mode being centered closer to the average cost for their white counterparts. Consequently, the differences in

liftime cost observed by race in hemodialysis patients may be somewhat overstated.

Two interesting trends identified by the above analysis are that hemodialysis tends to be more costly compared to peritoneal dialysis and that black patients tend to be more costly than white patients. However, the above analysis does not elucidate whether this difference is due to differences in survival times by modality or race, different baseline health measures on these patients, or whether e.g. hemodialysis is simply more costly. One way of standardizing lifetime costs to better compare hemodialysis versus peritoneal dialysis is to compute a cost-effectiveness ratio (CER). For a given mode of dialysis, the numerator of the CER represents some functional of lifetime cost; correspondingly, the denominator is a corresponding measure of treatment effectiveness (e.g. survival or quality-adjusted survival). See Siegel, Laska, and Meisner (1996) for a review of methods used in the economic appraisal of health care interventions.

We calculated CERs to evaluate the relative effectiveness of hemo- versus peritoneal dialysis for various subgroups of patients aged 66-70 years. We respectively computed the estimated mean lifetime cost and survival time (i.e., measured in months, and truncated at 4 years) using IPCW-HARE (see Table 7) and HARE; the results are shown in Table 8. Each CER thus represents "the average cost of dialysis per additional month of life with treatment." For the sake of comparison, we also computed and report CERs using median costs and survival in Table 8. Controlling for diabetic status, one tentative conclusion from this analysis is that peritoneal dialysis is in general less costly than hemodialysis for dialysis patients aged 66-70 years old. For example, the average additional cost per month of life for hemodialysis patients is \$1065 for diabetics and \$520 for nondiabetics, the differences in medians being somewhat lower. The trends observed by race in Table 7 are now less evident, due largely to the fact that black ESRD patients tend to live longer than white ESRD patients. Adjusting for survival differences, Table 8 in fact suggests that the median cost per month of life for treating white patients may exceed that for black patients, especially among hemodialysis patients. Finally, controlling for modality, another trend emerging from Table 8 is that diabetics tend to be more costly to treat than nondiabetics. For example, the additional average cost per month of life for diabetic patients is \$665 for hemodialysis patients and \$120 for nondiabetics. Importantly, this rudimentary cost-effectiveness analysis does not account for variability in the point estimates used in constructing these ratios, and one must be careful not to overinterpret these results. A more formal cost-benefit analysis might proceed from here with the computation of confidence intervals. Laska, Meisner, and Siegel (1997) give a review of methods in this area; a rather novel approach to evaluating CERs is described in Cook and Heyse (2000).

5 Discussion

The HARE model (1) is able to incorporate patient covariates while avoiding restrictive assumptions on their relationship to cost. We have established the utility of incorporating an IPCW weighting scheme into HARE for the purposes of dealing with censored lifetime medical cost data. By upweighting completely observed individuals, we are able to account for the informatively censored nature of the cost data, and at the same time avoid restrictive assumptions on the relationship between costs, survival, and covariates. This approach is quite useful in cases where understanding the cost structure is desirable for one or more covariate patterns, but less so for marginal analyses (i.e., V alone). Regarding the latter, reasonably efficient nonparametric estimators for functionals of the marginal cost distribution can be easily constructed; see, for example, van der Laan and Hubbard (1999), Bang and Tsiatis (2000b), or Strawderman (2000). However, one must impose further structure in order to model the conditional distribution $[V|Z]$. IPCW-HARE facilitates this with its highly data-adaptive capabilities, and our simulation studies show that it performs quite well in a variety of settings for fixed covariate patterns.

The addition of the IPCW weights into the HARE score equation necessitated substantial modification of the HARE fitting procedures. In large part, these changes were required in order to cope with the fact that one is no longer working within a maximum likelihood setting. From a practical point of view, these changes enhanced the robustness of fitting procedures, and in particular were found to substantially reduce the incidence of "overfitting". The decision to compute the matrix $\hat{Q}$ assuming $\hat{K}(\cdot|\cdot)$ is known was made primarily for practical reasons. Our simulation results indicate that the resulting estimators perform well. Limited comparisons suggest that the major consequence of estimating Q assuming $\hat{K}(\cdot|\cdot)$ is known rather than estimated is a slightly larger model. This appears to result from differences in the magnitude of the penalty term in the corresponding AICR statistic. Further research on the most appropriate criterion for model selection, as well as alternative methods for adaptively building models, is needed. Some recent work by Sin and White (1996) on more general Kullback-Leibler-based model selection criteria may prove useful here.

IPCW-HARE has a number of useful advantages over existing regression methods, many of which can be traced to the absence of restrictive parametric assumptions on the joint distribution of (V, T, Z). However, despite encouraging results, there are still some important practical limitations. Few of these are actually limitations of HARE itself, as we explain below. Major issues to consider when using this methodology include (i) the potential inefficiency of using only the data from complete cases; (ii) the unspecified dependence of the response variable V on T; (iii) the need to specify a model for the censoring mechanism; and, (iv) the failure to use the history of the cost process, if available.

We may handle the problem (i) by augmenting the IPCW-HARE estimating function to include additional information on censored subjects. Intuitively, adding any function of the censored cost data

with mean zero is a possible choice; the key is to select it so as to achieve a significant reduction in variability. Adjusting the IPCW-HARE estimating function in this manner would substantially increase the computational complexity of IPCW-HARE, and we will not attempt to explore these issues further here. In-depth discussions of such adjustments in related problems can be found in Rotnitzky, Robins and Scharfstein (1998) and Scharfstein, Robins, and Rotnitzky (1999).

The problem in (ii) stems from the unspecified dependence of V on T, which makes it difficult to assess whether the difference between, say, $E[V|Z = z_1]$ and $E[V|Z = z_2]$ is really due to differences in cost, differences in survival, or both. This is a direct result of the choice to model $[V|Z]$ instead of $[V,T|Z]$, a drawback not unique to this paper. Indeed, essentially all papers on this topic referenced here take this same point of view. In our example we used CERs to partially address this difficulty in interpretation. More formally, this problem might be addressed by modeling the conditional distribution $[V,T|Z] \equiv [V|T,Z][T|Z]$. In fact, including T as a covariate in the present implementation of IPCW-HARE appears to pose no fundamental difficulty with the theory, and consequently one could alleviate this difficulty in interpretation by adjusting cost comparisons to a common survival time. Moreover, HARE could be used to model $[T|Z]$, keeping with our goal of making as few parametric assumptions as possible on (V,T,Z).

The difficulty with (iii) is that, outside certain favorable settings (e.g., a randomized trial with only administrative censoring), the true censoring process is unknown. The importance of this rests in the fact that IPCW-HARE requires the correct model for censoring probabilities. Gross misspecification of these will certainly produce estimators whose properties are as questionable as those produced by standard survival techniques. However, the modeling burden must be placed somewhere. We have chosen to place this burden on the censoring process, a practice with significant value that has been established in numerous previous papers of Robins and his coauthors.

Finally, the problem with (iv) relates to the fact that our model only uses the cumulative cost at a patient's last follow-up or death time to build the distribution of lifetime cost. Models that incorporate aspects of a patient's entire cost history (e.g. monthly dialysis costs) will yield a richer, more complete picture of the lifetime cost distribution. Lin (2000b) and Bang and Tsiatis (2000a) propose handling this in much the same way, namely by considering a fixed partition of the time scale. Variations on IPCW-HARE could presumably be used in this setting as well. However, such approaches seem inefficient in general, since they do not utilize the possible correlation between cost increments across different time intervals. Methods for incorporating this information are a subject of ongoing research.

Acknowledgements

The authors would like to thank Drs. Fritz Port and Bob Wolfe of the USRDS and University of Michigan for useful discussions and financial support relevant to this project. They also acknowledge the assistance of Larry Agadoa, NIH Project Officer for the United States Renal System, in obtaining permission to use this USRDS data.

References

Andersen, P.K., Borgan, Ø., Gill, R.D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* New York: Springer-Verlag.

Bang, H. and Tsiatis, A.A. (2000a). Median Regression with Censored Cost Data. Unpublished.

Bang, H. and Tsiatis, A.A. (2000b). Estimating Medical Costs with Censored Data. *Biometrika*, 87, 329-344.

Cook, J.R. and Heyse, J.F. (2000). Use of an Angular Transformation for Ratio Estimation in Cost-Effectiveness Analysis. *Statistics in Medicine*, 19, 2989-3003.

Cook, R. and Lawless, J.F. (1997). Marginal Analysis of Recurrent Events and a Terminating Event. *Statistics in Medicine*, 16, 911-924.

Gelber, R.D., Gelman, R.S., and Goldhirsch, A. (1989). A Quality-of-Life-Oriented Endpoint for Comparing Therapies. *Biometrics*, 45, 781-795.

Heritier, S. and Ronchetti, E. (1994). Robust Bounded-Influence Tests in General Parametric Models. *Journal of the American Statistical Association*, 89, 897-904.

Hiriart-Urruty, J-B. and Lemaréchal, C. (1996). *Convex Analysis and Minimization Algorithms I.* Springer-Verlag, Berlin.

Kooperberg, C., Stone, C.J., and Truong, Y.K. (1995a). Hazard Regression. *Journal of the American Statistical Association*, 90, 78-94.

Kooperberg, C., Stone, C.J., and Truong, Y.K. (1995b). The L_2 Rate of Convergence for Hazard Regression. *Scandinavian Journal of Statistics*, 22, 143-157.

Laska, E., Meisner, M., and Siegel C. (1997). Statistical Inference for Cost-Effectiveness Ratios, *Health Economics*, 6, 229-242.

Lin, D.Y., Feuer, E.J., Etzioni, R., and Wax, Y. (1997). Estimating Medical Costs from Incomplete Follow-up Data. *Biometrics*, 53, 419-434.

Lin, D.Y. (2000a). Proportional Means Regression for Censored Medical Costs. *Biometrics*, 56, 775-778.

Lin, D.Y. (2000b). Linear Regression of Censored Medical Costs. *Biostatistics*, 1, 35-47.

Pollard, D. (1990). *Empirical Processes : Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 2. Institute of Mathematical Statistics, Hayward, Calif.

Robins, J.M. and Rotnitzky A. (1992). Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers. In *AIDS Epidemiology: Methodological Issues*, Ed. N. Jewell, K. Dietz, and V. Farewell, pp.297-331. Boston: Birkhäuser.

Ronchetti, E. (1997). Robustness Aspects of Model Choice. *Statistica Sinica*, 7, 327-338.

Rotnitzky, A., Robins, J., and Scharfstein, D. (1998). Semiparametric Regression for Repeated Outcomes with Nonignorable Nonresponse. *Journal of the American Statistical Association*, 93,1321-1339.

Scharfstein, D., Rotnitzky A., and Robins J. (1999). Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models. *Journal of the American Statistical Association*, 93, 1096-1120.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6, 461-464.

Siegel C., Laska, E., and Meisner, M. (1996). Statistical Methods for Cost-Effectiveness Analyses, *Controlled Clinical Trials*, 17, 387-406.

Sin, C-Y. and White, H. (1996). Information Criteria for Selecting Possibly Misspecified Parametric Models. *Journal of Econometrics*, 71, 207-225.

Strawderman, R.L. (2000). Estimating the Mean of an Increasing Stochastic Process at a Censored Stopping Time. *Journal of the American Statistical Association*, 95, 1192-1208.

Triplett, J.E. (1999). *Measuring the Prices of Medical Treatments.* Washington: Brookings Institution Press.

U.S. Renal Data System, USRDS 1999 Annual Data Report National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases Bethesda, MD, April 1999.

van der Laan, M.J. and Hubbard, A. (1999). Locally Efficient Estimation of the Quality-Adjusted Lifetime Distribution with Right-Censored Data and Covariates. *Biometrics*, 55, 530-536.

Zhao, H. and Tsiatis, A.A. (1997). A Consistent Estimator for the Distribution of Quality Adjusted Survival Time. *Biometrika*, 84, 339-348.

Table 1: Bias of Estimates for 1^{st} Cost Paradigm

		IPCW-HARE		Original HARE		Linear
	True Cost	PH=F	PH=T	PH=F	PH=T	Regression
Mean: Z=25	\$30,861	-\$468	-\$542	-\$1,746	-\$1,753	\$83
Median: Z=25	\$31,640	-\$50	-\$78	-\$1,376	-\$1,369	\$116
Mean: Z=35	\$34,396	-\$157	-\$198	-\$1,873	-\$1,873	-\$69
Median: Z=35	\$35,265	\$162	\$136	-\$1,560	-\$1,556	-\$15
Mean: Z=45	\$37,745	\$14	\$28	-\$1,841	-\$1,837	-\$34
Median: Z=45	\$38,726	\$205	\$200	-\$1,649	-\$1,649	\$18
Mean: Z=55	\$41,121	\$224	\$277	-\$1,795	-\$1,788	-\$26
Median: Z=55	\$42,206	\$290	\$301	-\$1,737	-\$1,739	\$32
Mean: Z=65	\$44,605	\$112	\$201	-\$1,819	-\$1,800	-\$126
Median: Z=65	\$45,773	\$0	\$33	-\$1,920	-\$1,926	-\$41

Caption Table 1: Simulation consists of 2000 data sets, each with n=200. We use the Cox proportional hazards model to estimate $K(.)$. "PH=" T: a proportional hazards model, F: covariate-by-cost interactions are allowed. The "true cost" is empirically obtained from a data set of size n =30,000 with the specified covariate pattern. The results in the "Linear Regression" column respectively correspond to the estimators of Lin (2000b) (rows labelled "mean") and Bang and Tsiatis (2000a) (rows labelled "median").

Table 2: Standard deviations under 1^{st} Cost Paradigm

		IPCW-HARE		Original HARE		Linear
	True Cost	PH=F	PH=T	PH=F	PH=T	Regression
Mean: Z=25	\$30,861	\$1,907	\$1,798	\$1,194	\$1,194	\$1,161
Median: Z=25	\$31,640	\$1,924	\$1,824	\$1,208	\$1,209	\$1,349
Mean: Z=35	\$34,396	\$1,265	\$1,303	\$957	\$959	\$794
Median: Z=35	\$35,265	\$1,240	\$1,278	\$945	\$944	\$919
Mean: Z=45	\$37,745	\$1,213	\$1,242	\$1,021	\$1,028	\$681
Median: Z=45	\$38,726	\$1,207	\$1,237	\$1,032	\$1,037	\$771
Mean: Z=55	\$41,121	\$1,335	\$1,384	\$1,172	\$1,177	\$920
Median: Z=55	\$42,206	\$1,357	\$1,405	\$1,229	\$1,242	\$1,034
Mean: Z=65	\$44,605	\$2,070	\$2,077	\$1,617	\$1,624	\$1,333
Median: Z=65	\$45,773	\$2,071	\$2,106	\$1,703	\$1,715	\$1,507

Caption Table 2: Empirical standard deviations over 2000 simulated datasets corresponding to Table 1.

Table 3: Bias of Estimates for 2^{nd} Cost Paradigm

		IPCW-HARE		Linear
	True Cost	PH=F	PH=T	Regression
Mean: Z=(15,0)	\$20,386	\$511	-\$1,053	-\$1,410
Median: Z=(15,0)	\$17,984	\$1,472	\$2,194	\$172
Mean: Z=(15,1)	\$25,452	-\$1,062	-\$2,089	-\$1,685
Median: Z=(15,1)	\$23,076	-\$338	-\$1,283	\$146
Mean: Z=(30,0)	\$38,614	-\$3	\$346	\$2,183
Median: Z=(30,0)	\$35,368	-\$795	-\$1,370	\$4,155
Mean: Z=(30,1)	\$43,695	-\$866	-\$1,527	\$1,894
Median: Z=(30,1)	\$40,595	\$230	-\$260	\$3,995
Mean: Z=(45,0)	\$64,305	\$244	\$817	-\$1,687
Median: Z=(45,0)	\$65,267	-\$189	\$501	-\$4,375
Mean: Z=(45,1)	\$69,143	\$67	-\$1,216	-\$1,733
Median: Z=(45,1)	\$70,206	-\$504	-\$1,677	-\$4,248

Caption Table 3: Simulation consists of 2000 data sets, each with n =200. We use the Cox proportional hazards model to estimate $K(.)$. "PH=" T: a proportional hazards model, F: covariate-by-cost interactions are allowed. The "true cost" is empirically obtained from a data set of size n =30,000 with the specified covariate pattern. The results in the "Linear Regression" column respectively correspond to the estimators of Lin (2000b) (rows labelled "mean") and Bang and Tsiatis (2000a) (rows labelled "median"). The censoring percentages for z_1 = 15, 30, and 45 are approximately 20%, 35%, and 45%, respectively.

Table 4: Standard Deviations for 2^{nd} Cost Paradigm

		IPCW-HARE		Linear
	True Cost	PH=F	PH=T	Regression
Mean: Z=(15,0)	\$20,386	\$2,090	\$2,159	\$3,196
Median: Z=(15,0)	\$17,984	\$1,801	\$1,787	\$2,601
Mean: Z=(15,1)	\$25,452	\$2,808	\$2,867	\$4,541
Median: Z=(15,1)	\$23,076	\$2,561	\$2,522	\$4,458
Mean: Z=(30,0)	\$38,614	\$3,311	\$3,422	\$3,012
Median: Z=(30,0)	\$35,368	\$3,912	\$3,962	\$4,096
Mean: Z=(30,1)	\$43,695	\$4,247	\$4,688	\$4,440
Median: Z=(30,1)	\$40,595	\$4,964	\$5,501	\$5,196
Mean: Z=(45,0)	\$64,305	\$6,472	\$7,455	\$5,789
Median: Z=(45,0)	\$65,267	\$7,920	\$8,460	\$7,799
Mean: Z=(45,1)	\$69,143	\$8,199	\$8,928	\$6,662
Median: Z=(45,1)	\$70,206	\$9,047	\$9,709	\$8,256

Caption Table 4: Empirical standard deviations over 2000 simulated datasets corresponding to Table 3.

Table 5: Bias of Estimates for 3^{rd} Cost Paradigm

	True Cost	IPCW-HARE PH=F	IPCW-HARE PH=T	Linear Regression
Mean: Z=(15,0)	\$1,309	-\$91	-\$63	\$1,422
Median: Z=(15,0)	\$1,432	-\$210	-\$203	\$988
Mean: Z=(15,1)	\$327	-\$13	-\$23	-\$1,579†
Median: Z=(15,1)	\$356	-\$43	-\$56	-\$658†
Mean: Z=(30,0)	\$4,840	-\$39	-\$147	\$462
Median: Z=(30,0)	\$5,173	-\$343	-\$543	-\$932
Mean: Z=(30,1)	\$1,212	-\$36	-\$30	\$107
Median: Z=(30,1)	\$1,295	-\$110	-\$134	\$224
Mean: Z=(45,0)	\$9,558	\$349	\$74	-\$1,684
Median: Z=(45,0)	\$9,930	-\$76	-\$304	-\$3,868
Mean: Z=(45,1)	\$2,388	\$19	\$66	\$1,503
Median: Z=(45,1)	\$2,481	-\$11	-\$59	\$858

Caption Table 5: Simulation consists of 2000 data sets, each with n =200. We use the Cox proportional hazards model to estimate $K(.)$. "PH=" T: a proportional hazards model, F: covariate-by-cost interactions are allowed. The "true cost" is empirically obtained from a data set of size n =30,000 with the specified covariate pattern. The results in the "Linear Regression" column respectively correspond to the estimators of Lin (2000b) (rows labelled "mean") and Bang and Tsiatis (2000a) (rows labelled "median"). The censoring percentages for (z_1, z_2) read from the top to the bottom of this table are respectively 49. 40, 44, 35, 34, and 28%. Bias estimates marked with a † correspond to a negative simulated average cost (mean or median).

Table 6: Standard Deviations for 3^{rd} Cost Paradigm

	True Cost	IPCW-HARE PH=F	IPCW-HARE PH=T	Linear Regression
Mean: Z=(15,0)	\$1,309	\$169	\$156	\$308
Median: Z=(15,0)	\$1,432	\$185	\$172	\$574
Mean: Z=(15,1)	\$327	\$44	\$40	\$389
Median: Z=(15,1)	\$356	\$49	\$45	\$420
Mean: Z=(30,0)	\$4,840	\$505	\$473	\$446
Median: Z=(30,0)	\$5,173	\$573	\$544	\$518
Mean: Z=(30,1)	\$1,212	\$130	\$120	\$176
Median: Z=(30,1)	\$1,295	\$145	\$130	\$147
Mean: Z=(45,0)	\$9,558	\$1,154	\$1,046	\$737
Median: Z=(45,0)	\$9,930	\$1,230	\$1,195	\$681
Mean: Z=(45,1)	\$2,388	\$270	\$259	\$389
Median: Z=(45,1)	\$2,481	\$314	\$277	\$352

Caption Table 6: Empirical standard deviations over 2000 simulated datasets corresponding to Table 5.

Table 7: Mean 4-Year Costs for Various USRDS Subgroups; 66-75 years old

	Race and Gender	Hemodialysis		Peritoneal Dialysis	
		Diabetic	Non-diabetic	Diabetic	Non-diabetic
66-70 years old	White Male	$142,900	$125,562	$105,369	$114,133
		(10,423)	(7,567)	(4,405)	(7,300)
	Black Male	$161,011	$147,759	$132,847	$132,751
		(16,930)	(16,322)	(9,867)	(14,069)
	White Female	$144,042	$124,891	$103,910	$112,262
		(11,080)	(7,158)	(3,933)	(7,418)
	Black Female	$160,843	$148,205	$134,043	$132,788
		(15,647)	(16,389)	(9,628)	(14,171)
71-75 years old	White Male	$120,977	$125,838	$105,430	$112,041
		(11,721)	(6,744)	(5,020)	(4,932)
	Black Male	$141,727	$149,886	$133,481	$132,918
		(15,999)	(15,178)	(10,800)	(13,503)
	White Female	$124,005	$125,527	$106,594	$113,072
		(10,050)	(6,005)	(4,402)	(5,654)
	Black Female	$142,219	$148,195	$133,614	$132,337
		(13,533)	(15,707)	(10,286)	(13,654)

Caption Table 7: Estimated mean costs obtained using IPCW-HARE. Numbers in parenthesis are standard errors based on 1000 bootstrap replications. Censoring percentages by column are approximately 20%, 24%, 17%, and 12%.

Table 8: CERs Based on Means and Medians for Various USRDS Subgroups 66-70 years old

	Race and Gender	Hemodialysis		Peritoneal Dialysis	
		Diabetic	Non-diabetic	Diabetic	Non-diabetic
Mean CER	White Male	$5,309	$4,396	$4,104	$3,921
	Black Male	$5,149	$4,590	$4,095	$4,062
	White Female	$5,179	$4,386	$4,025	$3,843
	Black Female	$4,980	$4,587	$4,132	$4,054
Median CER	White Male	$4,688	$4,002	$3,646	$3,481
	Black Male	$4,153	$3,305	$3,574	$3,214
	White Female	$4,338	$4,002	$3,646	$3,481
	Black Female	$3,915	$3,305	$3,574	$3,214

Caption Table 8: For each USRDS subgroup and covariate pattern, the mean cost-effectiveness ratio (CERs) is computed by taking the appropriate entry from Table 7 and dividing it by the corresponding average survival time, measured in months and truncated at 4 years. Median CERs are computed similarly (median costs and survival for these subgroups not shown).

6 Technical Appendix

Proposition 1 *Suppose that* $\widehat{K}(X|Z) > 0$ *with probability one. Further, assume* $P\{\Delta/\widehat{K}(X|Z) > 0\} > 0$ *and* $\sup_{u,z} \left|\widehat{K}(u|z) - K(u|z)\right| \xrightarrow{P} 0$. *Then, with probability tending to one, there exists a unique solution* $\widehat{\beta}$ *such that* $\widehat{\beta} \xrightarrow{P} \beta^*$, *where* $\beta^* \in I\!R^p$ *is the unique maximizer of the strictly concave function* $E[\ell_1(\beta)]$ *with* $\ell_1(\beta)$ *being given by (3).*

Proof of Proposition 1:

Andersen and Gill (1982) make use of concave function theory to prove consistency of the regression parameter under Cox's proportional hazards model. The HARE model, for a fixed set of basis functions, has a nice exponential family structure that allows direct use of Theorem II.1 and Corollary II.2 of Andersen and Gill (1982) for proving consistency of the p-dimensional solution vector $\widehat{\beta}$.

Let $\mathcal{B}$ be any open convex subset of $I\!R^p$ and define

$$H(\beta) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(X_i|Z_i)} \ell_i(\beta).$$

The exponential family structure of the HARE model (see e.g. Kooperberg *et al.*, 1995a, §3) ensures the likelihood contributions $\ell_i(\beta)$, $i = 1 \ldots n$ are concave functions of $\beta \in \mathcal{B}$. The assumptions of the proposition imply that, with probability tending to one, the weights $w_i = \Delta_i/\widehat{K}(X_i|Z_i), i = 1 \ldots n$ are nonnegative, finite, and such that $w_i > 0$ for at least some i. It follows that $H(\beta)$ forms a random sequence of concave functions in $\beta \in \mathcal{B}$ (cf. Hiriart-Urruty and Lemaréchal, 1996, Proposition 2.1.1). Moreover, for each $\beta \in \mathcal{B}$, $H(\beta) \xrightarrow{P} E[\ell_1(\beta)]$, where $E[\ell_1(\beta)]$ is itself concave. Hence, by Theorem II.1 of Andersen and Gill, $\sup_{\beta \in B} |H(\beta) - E[\ell_1(\beta)]| \xrightarrow{P} 0$ for any compact set $B \subset \mathcal{B}$.

The results of Kooperberg *et al.* (1995b) imply that $E[\ell_1(\beta)]$ has a unique maximum at $\beta = \beta^*$. Since our choice of $\mathcal{B}$ above is in fact arbitrary, let $\mathcal{B}$ be chosen such that $\beta^* \in B \subset \mathcal{B}$. Then, since $\widehat{\beta}$ maximizes $H(\beta)$, it follows by Corollary II.2 of Andersen and Gill (1982) that $\widehat{\beta} \xrightarrow{P} \beta^*$, proving the desired result. □

Proposition 2 *Suppose the hazard function for censoring follows the Cox proportional hazards model* $\lambda_C(u|z) = \lambda_{0C}(u) \exp\{\gamma_0^T z\}$. *For the data* $(X_i, 1 - \Delta_i, Z_i) i = 1 \ldots n$, *let* $\widehat{\gamma}$ *and* $\widehat{\Lambda}_{0C}(t)$ *be obtained in the usual manner, and assume the conditions of* §*VII.2.2 of Andersen* et al. *(1993) hold. Then,*

$$\sqrt{n} S_{ipcw}(\beta^*) = n^{-1/2} \sum_{i=1}^{n} \psi(Y_i, X_i, \Delta_i, Z_i, \beta^*, \gamma_0) + o_p(1)$$

where

$$\psi(Y_i, X_i, \Delta_i, Z_i, \beta^*, \gamma_0) = \frac{\Delta_i}{K(X_i|Z_i)} S_i(\beta^*) + \int_0^\infty g(u, V_i, T_i, Z_i, \beta^*, \gamma_0) dM_i^C(u), \tag{7}$$

$$M_i^C(u) = I\{X_i \leq u, \Delta_i = 0\} - \int_0^u I\{X_i \geq s\}\lambda_C(s|Z_i)ds,$$

and $g(u, V_i, T_i, Z_i, \beta^*, \gamma_0)$ *is given in (16).*

Proposition 3 *Consider the same setting as Proposition 2. Let* $S'_{ipcw}(\beta)$ *denote the derivative of* $S_{ipcw}(\beta)$ *with respect to* β*, and suppose that* $-S'_{ipcw}(\beta^*) \xrightarrow{P} M(\beta^*)$*, where* $M = M(\beta^*)$ *is nonsingular. Then,*

$$\sqrt{n}\left(\widehat{\beta} - \beta^*\right) \xrightarrow{d} N(0, M^{-1}QM^{-1})$$

for $Q = E[\psi(Y, X, \Delta, Z, \beta^*, \gamma_0)\psi^T(Y, X, \Delta, Z, \beta^*, \gamma_0)]$*. In particular, with* $v^{\otimes 2} = vv^T$ *for any vector* v*,*

$$Q = E\left[\frac{\{S_1(\beta^*)\}^{\otimes 2}}{K(T|Z)}\right] - \int_0^\infty \frac{\{w_0(u, \beta^*, \gamma_0)\}^{\otimes 2}}{s_0(u, \gamma_0)}\, d\Lambda_{0C}(u) - A\Sigma^{-1}A^T \tag{8}$$

where Σ^{-1} *is the asymptotic variance of* $\sqrt{n}(\widehat{\gamma} - \gamma_0)$ *and* $s_0(u, \gamma_0)$*,* $A = A(\beta^*, \gamma_0)$ *and* $w_0(u, \beta^*, \gamma_0)$ *are respectively defined in (13)-(15).*

REMARK: The matrix Q is derived directly from (7). Using the fact that $M_i^C(\cdot)$, $i = 1 \ldots n$ are orthogonal local square-integrable martingales with respect to a conveniently defined filtration (e.g., see Zhao and Tsiatis, 1997), the calculations involve simple and well-known variance and covariance identities for martingales. Estimation of Q may be carried out using appropriate approximations for each of the terms in (8). However, to ensure positive definiteness, it is preferable to compute Q via $n^{-1}\sum_i \widehat{\psi}_i\widehat{\psi}_i^T$, where $\widehat{\psi}_i$ is obtained from (7). In either case, the functions $w_0(u, \beta^*, \gamma_0)$ and matrix $A(\beta^*, \gamma_0)$ must be recomputed every time a basis function is added or deleted during the model building phase.

REMARK: The results in Propositions 2 and 3 in no way depend on $S_i(\beta)$ being the particular full data score derived under model (1). In fact, these asymptotic representations continue to apply to any mean zero function of the full data having finite variance, provided $K(\cdot|\cdot)$ is estimated via a Cox model. Hence, for example, our results contain those of Lin (2000b) for a single time interval as a special case.

Proof of Proposition 2:

We suppose Z is $q \times 1$ and β is $p \times 1$. Also, for any vector v, we let $v^{\otimes 0} = 1$, $v^{\otimes 1} = v$, and $v^{\otimes 2} = vv^T$. We may write

$$\frac{\Delta_i}{\widehat{K}(X_i|Z_i)} = \frac{\Delta_i}{K(X_i|Z_i)} - \frac{\Delta_i}{\widehat{K}(X_i|Z_i)}\left(\frac{\widehat{K}(X_i|Z_i)}{K(X_i|Z_i)} - 1\right). \tag{9}$$

As pointed out in Andersen *et al.* (1993, §VII.2.3), there are a variety of different, yet asymptotically equivalent, ways to estimate $K(\cdot|z)$ based on $\widehat{\gamma}$ and $\widehat{\Lambda}_{0C}(t)$. Let us assume $K(\cdot|z)$ is estimated as in (7.2.34) of Andersen *et al.* (1993); then, employing the Duhamel equation (cf. Andersen *et al.*, 1993, eqn. 2.6.5),

$$\frac{\widehat{K}(X_i|Z_i)}{K(X_i|Z_i)} - 1 = -\int_0^\infty R_i(u)\frac{\widehat{K}(u^-|Z_i)}{K(u|Z_i)}\, dL_i(u) \tag{10}$$

where $L_i(u) = \widehat{\Lambda}_C(u|Z_i) - \Lambda_C(u|Z_i)$, $\widehat{\Lambda}_C(s|Z_i) = \widehat{\Lambda}_{0C}(s)\exp\{\widehat{\gamma}^T Z_i\}$, $\Lambda_C(s|Z_i) = \Lambda_{0C}(s)\exp\{\gamma_0^T Z_i\}$, and $R_i(u) = I\{X_i \geq u\}$. Using (9) and (10), we thus find

$$S_{ipcw}(\beta^*) = n^{-1}\sum_{i=1} S_i(\beta^*)\left[\frac{\Delta_i}{K(X_i|Z_i)} + \frac{\Delta_i}{\widehat{K}(X_i|Z_i)}\int_0^\infty R_i(u)\frac{\widehat{K}(u^-|Z_i)}{K(u|Z_i)}\, dL_i(u)\right]. \tag{11}$$

Using results in Andersen *et al.* (1993, pp. 503-506), it is not difficult to establish that, for a given $Z = z^*$ and $t > 0$, the following expansion holds for $\widehat{\Lambda}_C(t|z^*) - \Lambda_C(t|z^*)$:

$$\widehat{\Lambda}_C(t|z^*) - \Lambda_C(t|z^*) = J^T(t,\gamma_0,z^*)(\widehat{\gamma}-\gamma_0) + n^{-1}\exp\{\gamma_0^T z^*\}\int_0^t \frac{d\bar{M}^C(u)}{s_0(u,\gamma_0)} + o_p(n^{-1/2}), \tag{12}$$

where

$$J(t,\gamma_0,z^*) = \exp\{\gamma_0^T z^*\}\int_0^t \left(z^* - \frac{s_1(u,\gamma_0)}{s_0(u,\gamma_0)}\right) d\Lambda_{0C}(u),$$

$$s_k(u,\gamma_0) = E[Z^{\otimes k} I\{X \geq u\}\exp\{\gamma_0^T Z\}],\ k = 0,1 \tag{13}$$

and $\bar{M}^C(u) = \sum_{j=1}^n M_j^C(u)$ for $M_j^C(u) = I\{X_j \leq u, \Delta_j = 0\} - \int_0^u Y_j(s)d\Lambda_C(s|Z_j)$. Substituting (12) into (11) (i.e., with $z^* = Z_i$) and rearranging terms,

$$S_{ipcw}(\beta^*) = n^{-1}\sum_{i=1}^n \frac{\Delta_i S_i(\beta^*)}{K(X_i|Z_i)} + \widehat{A}(\widehat{\gamma}-\gamma_0) + n^{-1}\sum_{i=1}^n \int_0^\infty \left[n^{-1}\sum_{j=1}^n \xi_j(u)S_j(\beta^*)\right]\frac{dM_i^C(u)}{s_0(u,\gamma_0)} + o_p(n^{-1/2})$$

where

$$\widehat{A}^{p\times q} = \int_0^\infty \left[n^{-1}\sum_{i=1}^n \xi_i(u)S_i(\beta^*)\left(Z_i - \frac{s_1(u,\gamma_0)}{s_0(u,\gamma_0)}\right)^T\right] d\Lambda_{0C}(u).$$

and

$$\xi_i(u) = \frac{\Delta_i}{K(X_i|Z_i)}R_i(u)\exp\{\gamma_0^T Z_i\}.$$

Let $b(V,T,Z)$ be any bounded function of the full data vector (V,T,Z). Then, since $\Delta I\{X \geq u\} = \Delta I\{T \geq u\}$, it is easy to show that

$$E[\xi(u)b(V,T,Z)] = E[I\{T \geq u\}\exp\{\gamma_0^T Z\}b(V,T,Z)],$$

and hence that

$$\sup_{u\geq 0}\left|n^{-1}\sum_{i=1}^n \xi_i(u)b(V_i,T_i,Z_i) - E[I\{T \geq u\}\exp\{\gamma_0^T Z\}b(V,T,Z)]\right| \xrightarrow{P} 0$$

(e.g., Pollard, 1990, Theorem 8.3). Thus,

$$S_{ipcw}(\beta^*) = n^{-1}\sum_{i=1}^n \left[\frac{\Delta_i}{K(X_i|Z_i)}S_i(\beta^*) + \int_0^\infty \frac{w_0(u,\beta^*,\gamma_0)}{s_0(u,\gamma_0)}dM_i^C(u)\right] + A(\widehat{\gamma}-\gamma_0) + o_p(n^{-1/2}),$$

where $A = A^{p\times q}(\beta^*,\gamma_0)$,

$$A^{p\times q}(\beta^*,\gamma_0) = \int_0^\infty \left(w_1(u,\beta^*,\gamma_0) - w_0(u,\beta^*,\gamma_0)\frac{s_1^T(u,\gamma_0)}{s_0(u,\gamma_0)}\right) d\Lambda_{0C}(u), \tag{14}$$

and

$$w_k(u, \beta^*, \gamma_0) = E[S_1(\beta^*) Z^{\otimes k} I\{T \geq u\} \exp\{\gamma_0^T Z\}], \ k = 0, 1. \tag{15}$$

Finally, since

$$\widehat{\gamma} - \gamma_0 = n^{-1} \Sigma^{-1}(\gamma_0) \sum_{i=1}^{n} \int_0^\infty \left(Z_i - \frac{s_1(u, \gamma_0)}{s_0(u, \gamma_0)} \right) dM_i^C(u) + o_p(n^{-1/2})$$

(cf. Andersen *et al.*, 1993, §VII.2.2), it follows that

$$\sqrt{n} S_{ipcw}(\beta^*) = n^{-1/2} \sum_{i=1}^{n} \psi(Y_i, X_i, \Delta_i, Z_i, \beta^*, \gamma_0) + o_p(1)$$

where

$$\psi(Y_i, X_i, \Delta_i, Z_i, \beta^*, \gamma_0) = \frac{\Delta_i}{K(X_i|Z_i)} S_i(\beta^*) + \int_0^\infty g(u, V_i, T_i, Z_i, \beta^*, \gamma_0) dM_i^C(u)$$

with

$$g(u, V_i, T_i, Z_i, \beta^*, \gamma_0) = \frac{w_0(u, \beta^*, \gamma_0)}{s_0(u, \gamma_0)} + A(\beta^*, \gamma_0) \Sigma^{-1}(\gamma_0) \left(Z_i - \frac{s_1(u, \gamma_0)}{s_0(u, \gamma_0)} \right). \tag{16}$$

Noting that

$$\frac{\Delta_i}{K(X_i|Z_i)} = 1 - \int_0^\infty \frac{dM_i^C(u)}{K(u|Z_i)}$$

(e.g., Strawderman, 2000, Lemma 1), one may rewrite the contribution of individual i as

$$\psi(Y_i, X_i, \Delta_i, Z_i, \beta^*, \gamma_0) = S_i(\beta^*) - \int_0^\infty \frac{S_i(\beta^*)}{K(u|Z_i)} dM_i^C(u) + \int_0^\infty g(u, V_i, T_i, Z_i, \beta^*, \gamma_0) dM_i^C(u),$$

a form useful for variance calculations (cf. Zhao and Tsiatis, 1997).