3-2-2007

# SEMIPARAMETRIC BIVARIATE QUANTILE-QUANTILE REGRESSION FOR ANALYZING SEMI-COMPETING RISKS DATA

Daniel O. Scharfstein
*Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*, dscharf@jhsph.edu

James M. Robins
*Department of Epidemiology and Biostatistics, Harvard School of Public Health*

Mark van der Laan
*Department of Biostatistics and Statistics, University of California, Berkeley*

# Semiparametric Bivariate Quantile-Quantile Regression Model for Analyzing Semi-Competing Risks Data

Daniel O. Scharfstein, James M. Robins, and Mark van der Laan *

March 2, 2007

**Abstract**

In this paper, we consider estimation of the effect of a randomized treatment on time to disease progression and death, possibly adjusting for high-dimensional baseline prognostic factors. We assume that patients may or may not have a specific type of disease progression prior to death and those who have this endpoint are followed for their survival information. Progression and survival may also be censored due to loss to follow-up or study termination. We posit a semi-parametric bivariate quantile-quantile regression failure time model and show how to construct estimators of the regression parameters. The causal interpretation of the parameters depends on non-identifiable assumptions. We discuss two assumptions: the first applies to situations where it is reasonable to view disease progression as well defined after death and the second applies to situations where such a view is unreasonable. We conduct a simulation study and analyze data from a randomized trial for the treatment of brain cancer.

KEYWORDS: Brain Tumors; Causal Inference; Censoring by Death; Estimating Equations



*Daniel O. Scharfstein is an Associate Professor of Biostatistics at Johns Hopkins Bloomberg School of Public Health. James M. Robins is Professor of Epidemiology and Biostatistics at Harvard School of Public Health. Mark van der Laan is Professor of Biostatistics and Statistics, University of California, Berekeley

# 1 Introduction

In randomized trials to evaluate treatment of life-threatening diseases, subjects are often monitored for specific types of disease progression and survival. In such studies, a progression endpoint may be pre-empted by death or censored due to loss to follow-up or study termination. Subjects who experience a progression event are also followed for survival, which may be censored. Data of this form has been labeled semi-competing risks data (Fine, Jiang, and Chappell, 2001). Our motivating example of a study which produces such data is a randomized trial for the treatment of malignant brain tumors. In the study, one of the important progression endpoints was based on deterioration, from baseline, of the cerebellum. An important feature of this endpoint is that it is biologically plasuible that a subject could die without cerebellar deterioration.

Robins (1995a, 1995b) and Lin *et al.* (1996) introduced a semi-parametric bivariate location-shift model to describe the joint effect of treatment on progression and survival in two-arm randomized studies. This model can be viewed as a bivariate linear regression with log of the survival and progression endpoints as the response, correlated error terms, and treatment indicator as the sole covariate. The regression parameter for survival is estimated using the estimating function technique of Louis (1981), Wei and Gail (1983) and Tsiatis (1990). Working on the residual death time scale, the estimating function is based on the comparison of the observed treatment indicator at each death time to the expected value of the treatment indicator among subjects at risk at that time. Lin *et al.* (1996) noticed that application of this technique to the progression endpoint leads to a biased estimating function because (1) progression can be pre-empted by death and (2) death and progression are correlated events. To solve this problem, Lin *et al.* (1996) used an idea originally introduced by Robins and Rotnitzky (1992, Appendix 4) and corrected the bias of the estimating function by artificially censoring the progression residuals. This approach works well for a low-dimensional covariate, such as a treatment indicator, but excessive artificial censoring can occur when the covariates are high-dimensional. To address this

latter problem, Peng and Fine (2006) introduced a new artificial censoring technique based on pairwise ranking. While this technique represents an improvement over the approach of Lin et al. (1996), it can, contrary to the suggestion of those authors, be less efficient than the alternative method introduced in this paper. There has been increased use and extension of the bivariate linear regression model in the analysis of recurrent event data and repeated measures data subject to informative drop-out (see, for example, Chang (2000), Joffe (2001), Ghosh and Lin (2003), Lin and Ying (2003), and Matsui (2004)). All these approaches use the artificial censoring technique of Lin *et al.* (1996) and, as a result, have difficulty dealing with high-dimensional covariates. In this paper, we introduce an alternative estimating equation that does not rely on the artificial censoring technique.

The paper is organized as follows. In Section 2, we introduce the study that motivated this research. In Section 3, we introduce the data structure, notation, and the generalized bivariate quantile-quantile regression model and its special case the bivariate linear regression model. Section 4 discusses issues of identifiability and Section 5 discusses the causal interpretation of the treatment effect regression parameters. Section 6 is devoted to estimation, while Section 7 discusses large sample theory. In Section 8, we present the results of a simulation study and Section 9 provides an analysis of the brain tumor trial. The final section is devoted to a discussion. To ease the flow of presentation, proofs of lemmas and theorems are placed in the Appendix. For readers who are less concerned with technical details, Section 7 and the Appendix can be skipped without loss of continuity.

# 2 Brain Tumor Study

## 2.1 Gliomas

Gliomas are primary brain tumors. Treatment for glioma involves maximal surgical resection of the tumor, followed by radiotherapy and chemotherapy. Prior to 1990, adjuvantchemother-

apy for brain tumors was limited because of the difficulty in achieving adequate exposure to the tumor site without causing systemic toxicity. In a phase 1 study, Brem *et al.* (1991) showed this diificulty could be overcome by implanting at surgery local to the tumor, a biodegradable polymers impreganted with the drug carmustine (BCNU), resulting in prolonged local exposure to BCNU with minimal systemic exposure.

## 2.2 Glioma Trial

The Phase I study was followed by a randomized trial, in which 222 patients with recurrent malignant brain tumors scheduled for tumor resection were randomly assigned to receive surgically implanted biodegradable polymer discs with or without 3.85% of carmustine (Brem *et al.* 1995). Patients were included in the study if they had a single focus of tumor in the cerebrum, a Karnofsky score greater than 60, completed radiation therapy, not taken nitrosoureas within 6 weeks of enrollment, and did not have systemic chemotherapy within 4 weeks of enrollment.

All subjects were followed for 1 year and were clinically and radiologically assessed at baseline and at least once every two months, thereafter. In particular, subjects were evaluated on 11 pre-specified neuroperformance measures, including an examination of cerebellar function. In the study, the primary endpoint was survival and secondary endpoints included neurological progression. In our analysis, we focused on the cerebellar examination measure as our secondary endpoint. Of the 219 subjects with complete baseline information, 204 were observed to die, 100 subjects were observed to progress on the cerebellar examination prior to death, and of the 15 subjects who did not die, 4 were observed to have cerebellar progression.

Besides treatment assignment, other important baseline prognostic factors included age, race, Karnofsky performance score, local vs. whole brain radiation, "active" vs. "quiescent" tumors, percent of tumor resection, previous use of nitrosoureas, and tumor histology

(glioblastoma, anaplastic astrocytoma, oligodendroglioma, or other) at implantation, For the 219 subjects with complete prognostic inofrmation, Table 1 compares summary statistics for each of these factors, stratified by treatment group. In Section 7, we will precisely define

### 2.3 Scientific Objective

The main objective is to evaluate the joint causal effect of treatment both on time to death and on time to cerebellar progression, while adjusting for apriori-defined baseline prognostic factors. The proper definition, much less estimation, of the effect of treatment on cerebellar progression is subtle and will be discussed briefly in the following section after we introduce the necessary notation. An extensive discussion is deferred to Section 5.

## 3 Data and Model

### 3.1 Data

Let $Z = (V, W')'$, where $V$ denotes a binary treatment indicator and $W$ is a $q$-dimensional bounded, random vector of additional regressors with a mixture of possibly discrete and continuous components. We assume that the $k$th component of $Z$ has support of $[l_k, u_k]$, where $-\infty < l_k < u_k < \infty$. Since the first component of $Z$ is binary, $l_0 = 0$ and $u_0 = 1$. Let $Y^0$ denote the logarithm of the time to death, $X^0$ denote the logarithm of time of disease progression, and $C$ denote the logarithm of the censoring time due to random loss to follow-up. $X^0$ and $Y^0$ may be correlated. Let $X = X^0 \wedge Y^0 \wedge C$, $\delta = I(X^0 \leq Y^0 \wedge C)$, $Y = Y^0 \wedge C$, and $\xi = I(Y^0 \leq C)$. . The observed data for an individual are $O = (X, Y, Z, \delta, \xi)$. Note that $Y^0$ censors $X^0$ but not vice versa.

## 3.2 Model

We specify our model without regard to censoring. We will address the issue of censoring in the identifiability and estimation sections to follow.

We will assume, until Section 5, that $X^0$ is well-defined for subjects with $X^0 \not< Y^0$. When $Y^0$ denotes time to death, this assumption has been criticized by Kalbfleisch and Prentice (1980), who argue that $X^0$ should be considered as undefined. It is of interest to consider estimation of causal effects under this latter assumption because

(i) in many studies, $Y^0$ does not denote time to death and the methodology developed in this paper is applicable to such studies, e.g,, in a short-term study comparing the effects of two antidepressant on time to first clinical improvement ($X^0$) and to non-compliance ($Y^0$), where scientific interest lies in regarding $X^0$ as censored by $Y^0$ but not vice-versa.

(ii) even when $Y^0$ is death time, settings may exist in which it is reasonable to regard $X^0$ as well defined even when $X^0 \not< Y^0$. As in Robins and Greenland (2000), consider a study in which a cohort of young children in a developing country are followed for, say 5 years, for the development of abnormal blood pressure. Since children are highly unlikely to die from hypertension, it might be reasonable to imagine the time that they would have developed abnormal blood pressure had their death been prevented. It would not be reasonable if, instead, we had followed a cohort of adults.

In Section 5.2.2, we provide conditions under which our method delivers consistent estimates of the effect of treatment on $X^0$, when $X^0$ is considered undefined for subjects $X^0 \not< Y^0$.

We now define the *bivariate quantile-quantile regression* model. The model is expressed as a quantile-quantile mapping between conditional distributions. Specifically, the

model assumes that

$$S_{Y^0}(t|V=1,W=w) = S_{Y^0}(r(t,w;\eta_{0a})|V=0,W=w) \tag{1}$$

$$S_{Y^0}(t|V=0,W=w) = S_{Y^0}(r^*(t,w;\eta_{0b})|V=0,W=0) \tag{2}$$

and

$$S_{X^0}(t|V=1,Y^0=r^{-1}(u,w;\eta_{0a}),W=w) = S_{X^0}(q(t,u,w;\theta_{0a})|V=0,Y^0=u,W=w) \tag{3}$$

$$S_{X^0}(t|V=0,Y^0=r^{*-1}(u,w;\eta_{0b}),W=w) = S_{X^0}(q^*(t,u,w;\theta_{0b})|V=0,Y^0=u,W=0) \tag{4}$$

where $r(t,w;\eta_a)$, $r^*(t,w;\eta_b)$, $q(t,u,w;\theta_a)$, and $q^*(t,u,w;\theta_b)$ are known increasing functions of $t$ with well defined inverses with respect to their first arguments, $r^*(t,0;\eta_{0b}) = t$, $q^*(t,u,0;\theta_{0b}) = t$,

- $r(t,w;\eta_{0a})$ is the quantile-quantile mapping function between the distribution of time to death among treated subjects with covariates $w$ and the distribution of time to death among untreated subjects also with covariates $w$ (see Figure 1),
- $r^*(t,w;\eta_{0b})$ is the quantile-quantile mapping function between the distribution of time to death among untreated subjects with covariates $w$ and the distribution of time to death among untreated subjects with covariates $W=0$,
- $q(t,u,w;\theta_{0a})$ is the quantile-quantile mapping between the distribution of time to progression among treated subjects with covariates $w$ who die at $r^{-1}(u,w;\eta_{0a})$ and the distribution of time to progression among untreated subjects with covariates $w$ who die at $u$ (see Figure 2), and
- $q^*(t,u,w;\theta_{0b})$ is the quantile-quantile mapping between the distribution of time to progression among untreated subjects with covariates $w$ who die at $r^{*-1}(u,w;\eta_{0b})$ and

the distribution of time to progression among untreated subjects with covariates $w$ who die at $u$ with covariates $W = 0$.

Define $\eta_0 = (\eta'_{0a}, \eta'_{0b})'$ and $\theta_0 = (\theta'_{0a}, \theta'_{0b})'$ as the true parameter vectors with $\eta_{0a}$, $\eta_{0b}$, $\theta_{0a}$, and $\theta_{0b}$ variation independent. Note that the right hand side of (3) is conditioned on untreated subjects who die at time $u$ with covariates $w$, whereas the right hand side is conditioned on treated subjects with covariates $w$ who die at the same percentile of the conditional death time distribution as those who are untreated (since, from (1), $S_{Y^0}(u|V = 0, W = w) = S_{Y^0}(r^{-1}(u, w; \eta_0)|V = 1, W = w)$).

The above model can be re-written as follows:

$$S_{Y^0}(t|V = 1, W = w) = S_{Y^0}(m(t, w; \eta_0)|V = 0, W = 0) \tag{5}$$

where $m(t, w; \eta_0) = r^*(r(t, w; \eta_{0a}), w; \eta_{0b})$ and

$$S_{X^0}(t|V = 1, Y^0 = m^{-1}(u, w; \eta_0), W = w) = S_{X^0}(l(t, u, w; \eta_{0b}, \theta_0)|V = 0, Y^0 = u, W = w) \tag{6}$$

where $l(t, u, w; \eta_{0b}, \theta_0) = q^*(q(t, r^{*-1}(u, w; \eta_{0b}), w; \theta_{0a}), u, w; \theta_{0b})$

Define the following random variables

$$\tilde{R}(\eta_{0a}) = Vr(Y^0, W; \eta_{0a}) + (1 - V)Y^0 \tag{7}$$

$$\begin{aligned}\tilde{M}(\eta_0) &= Vm(Y^0, W; \eta_0) + (1 - V)r^*(Y^0, W; \eta_{0b}) \\ &= \tilde{M}(Y^0, Z; \eta_0)\end{aligned} \tag{8}$$

$$\tilde{Q}(\eta_{0a}, \theta_{0a}) = Vq(X^0, \tilde{R}(\eta_{0a}), W; \theta_{0a}) + (1 - V)X^0 \tag{9}$$

$$\begin{aligned}\tilde{L}(\eta_0, \theta_0) &= Vl(X^0, \tilde{M}(\eta_0), W; \eta_{0b}, \theta_0) + (1 - V)q^*(X^0, \tilde{M}(\eta_0), W; \theta_{0b}) \\ &= \tilde{L}(X^0, \tilde{M}(Y^0, Z; \eta_0), Z; \theta_0)\end{aligned} \tag{10}$$

Under(1)-(6), we can prove the following lemma (see Appendix):

**Lemma 1.** *(a) $\tilde{R}(\eta_{0a})$ is independent of $V$ given $W$; (b) $\tilde{M}(\eta_0)$ is independent of $Z$; (c) $\tilde{Q}(\eta_{0a}, \theta_{0a})$ is independent of $V$ given $W$ and $\tilde{R}(\eta_{0a})$; (d) $\tilde{L}(\eta_0, \theta_0)$ is independent of $Z$ given $\tilde{M}(\eta_0)$; (e) $(\tilde{M}(\eta_0), \tilde{L}(\eta_0, \theta_0))$ is independent of $Z$.*

To link this model to other model forms, consider the following model:

$$Y^0 = V g_1(W; \eta_{0a}) + g_2(W; \eta_{0b}) + \nu \tag{11}$$

$$X^0 = V h_1(W, g_2(W; \eta_{0b}) + \nu; \theta_{0a}) + h_2(W, \nu; \theta_{0b}) + \epsilon \tag{12}$$

where $g_1$, $g_2$, $h_1$ and $h_2$ are specified functions of their arguments, $g_2(0; \eta_{0b}) = 0$, $h_2(0, \nu; \theta_{0b}) = 0$ and $(\nu, \epsilon)$ are joint independent of $Z$. No restrictions are placed on the joint distribution of $\nu$ and $\epsilon$. This model is a special case of the bivariate quantile-quantile regression model with

$$\begin{aligned}
r(t, w; \eta_{0a}) &= t - g_1(w; \eta_{0a}) \\
r^*(t, w; \eta_{0b}) &= t - g_2(w; \eta_{0b}) \\
m(t, w; \eta_0) &= t - g_1(w; \eta_{0a}) - g_2(w; \eta_{0b}) \\
q(t, u, w; \theta_{0a}) &= t - h_1(w, u; \theta_{0a}) \\
q^*(t, u, w; \theta_{0b}) &= t - h_2(w, u; \theta_{0b}) \\
l(t, u, w; \theta_0, \eta_{0b}) &= t - h_1(w, g_2(w; \eta_{0b}) + u; \theta_{0a}) - h_2(w, u; \theta_{0b}) \\
\tilde{M}(\eta_0) &= \nu \\
\tilde{L}(\eta_0, \theta_0) &= \epsilon
\end{aligned}$$

The following special case of the above model with

$$g_1(W;\eta_{0a}) = \eta_{0a}$$
$$g_2(W;\eta_{0b}) = W'\eta_{0b}$$
$$h_1(W,u;\theta_{0a}) = \theta_{0a}$$
$$h_2(W,u;\theta_{0b}) = W'\theta_{0b}$$

is the *bivariate linear regression* model:

$$Y^0 = V\eta_{0a} + W'\eta_{0b} + \nu \tag{13}$$
$$X^0 = V\theta_{0b} + W'\theta_{0b} + \epsilon \tag{14}$$

Lin, Wei, and Robins (1996) considered the special case without covariates $W$. They referred to their model as the *bivariate location-shift* model. Robins 1995 (a,b) and Peng and Fine (2006) considered the case with covariates.

Henceforth, we again allow censoring by administrative censoring or loss to follow-up $C$. Further, we assume that $C$ and $Z$ are jointly independent of $\tilde{M}(\eta_0)$ and $\tilde{L}(\eta_0,\theta_0)$.

# 4 Identifiability of $\eta_0$ and $\theta_0$

## 4.1 Identifiability of $\eta_0$

The parameter $\eta_0$ in the bivariate quantile-quantile regression model is identifiable because the conditional survivor function, $S_{Y^0}(\cdot|V=v,W=w)$, for $v=0,1$ and all $w$, is identifiable. Under the independent censoring assumption and the fact that, given $V$ and $W$, there is a

one-to-one mapping between $Y^0$ and $\tilde{M}(\eta_0)$, and

$$\lambda_{Y^0}(y|V=v,W=w) = \lambda_{Y^0}(y|V=v,W=w,C\geq y) \tag{15}$$

where $\lambda_{Y^0}(y|\cdot) = \lim_{dy\to 0} P[y \leq Y^0 < y+dy|Y^0 \geq y,\cdot]/dy$ is the hazard for $Y^0$ given $\cdot$. The right hand side of (15) is the cause-specific hazard for death and is identifiable from the distribution of the observed data. The conditional survivor function is identifiable since it be written in terms of the cause-specific hazard as follows:

$$S_{Y^0}(t|V=v,W=w) = \exp\left(-\int_{-\infty}^{t} \lambda_{Y^0}(y|V=v,W=w,C\geq y)dy\right),$$

Thus,

$$\begin{aligned}
r(t,w;\eta_{0a}) &= S^{-1}_{Y^0}(S_{Y^0}(t|V=1,W=w)|V=0,W=w)\\
r^*(t,w;\eta_{0b}) &= S^{-1}_{Y^0}(S_{Y^0}(t|V=0,W=w)|V=0,W=0)\\
m(t,w;\eta_0) &= S^{-1}_{Y^0}(S_{Y^0}(t|V=1,W=w)|V=0,W=0)
\end{aligned}$$

We assume that the conditional distribution $Y^0$ given $V=v$ and $W=w$ has positive support on $(-\infty,\tau]$ and $P[C>\tau|V=v,W=w]>0$, for some specified $-\infty<\tau<\infty$ and for all $v$ and $w$. So, $r(t,w;\eta_{0a})$, $r^*(t,w;\eta_{0b})$, and $m(t,w;\eta_0)$ will be identifiable for $t\in(-\infty,\tau]$ for all $w$. We only define these functions for $t\in(-\infty,\tau]$.

## 4.2 Identifiability of $\theta_0$

The identification argument for $\theta_0$ in bivariate quantile-quantile regression model is more subtle. The conditional survivor function $S_{X^0}(\cdot|V=v,Y^0=y,W=w)$ is only identifiable on the support $(-\infty,y]$. Under the independent censoring assumption and the fact that,

given $V$ and $W$ and $Y^0$, there is a one-to-one mapping between $X^0$ and $\tilde{L}(\eta_0, \theta_0)$,

$$\lambda_{X^0}(x|V=v, Y^0=y, W=w) = \lambda_{X^0}(x|V=v, Y^0=y, W=w, C \geq y) \tag{16}$$

where $\lambda_{X^0}(x|\cdot) = \lim_{dx \to 0} P[x \leq X^0 < x + dx | X^0 \geq x, \cdot]$ is the hazard for $X^0$ given $\cdot$. The right hand side of (16) is identifiable from the observed data distribution for $x \leq y$. Thus, the conditional survivor function is identifiable on $(-\infty, y]$ since, we can write it as

$$S_{X^0}(t|V=v, Y^0=y, W=w) = \exp\left( - \int_{-\infty}^{t} \lambda_{X^0}(x|V=v, Y^0=y, W=w, C \geq y) dx \right),$$

We assume that the conditional distributon of $X^0$ given $V=v$, $Y^0=y$, and $W=w$ has positive density on $(-\infty, y]$. Since the right and left hand sides of (3,4,6) are conditioned on different times of death, the quantile-quantile mapping functions $q(t,u,w;\theta_{0a})$, $q^*(t,u,w;\theta_{0b})$, and $l(t,u,w;\eta_{0b},\theta_0)$ are only identified, respectively, on the sets

$$\begin{aligned}
\mathcal{A}_q &= \{(t,u,w) : t \leq \min\{r^{-1}(u,w;\eta_{0a}), q(u,u,w;\theta_{0a})\}\} \\
\mathcal{A}_{q^*} &= \{(t,u,w) : t \leq \min\{r^{*-1}(u,w;\eta_{0a}), q^*(u,u,w;\theta_{0b})\}\} \\
\mathcal{A}_l &= \{(t,u,w) : t \leq \min\{m^{-1}(u,w;\eta_0), l(u,u,w;\eta_{0b},\theta_0)\}\}
\end{aligned}$$

Since, by definition,

$$\begin{aligned}
q(t,u,w:\theta_{0a}) &= S^{-1}_{X^0}(S_{X^0}(t|V=1, Y^0=r^{-1}(u,w;\eta_{0a}), W=w)|V=0, Y^0=u, W=w) \\
q^*(t,u,w:\theta_{0b}) &= S^{-1}_{X^0}(S_{X^0}(t|V=0, Y^0=r^{*-1}(u,w;\eta_{0b}), W=w)|V=0, Y^0=u, W=0) \\
l(t,u,w:\eta_{0a},\theta_0) &= S^{-1}_{X^0}(S_{X^0}(t|V=1, Y^0=m^{-1}(u,w;\eta_0), W=w)|V=0, Y^0=u, W=w)
\end{aligned}$$

we can see that these functions may not be identified off their respective sets above. For example, if we had specified the model

$$q(t,u,w;\theta_a) = (t - \theta_{a1})\, I\left((t,u,w) \in A_q\right) + (t - \theta_{a2})\, I\left((t,u,w) \notin A_q\right)$$

where $\theta_a = (\theta_{a1}, \theta_{a2})$, $\theta_{a1}$ and $\theta_{a2}$ variation independent and $\theta_{0a} = (\theta_{0a1}, \theta_{0a2})$, then $\theta_{0a1}$ would be identified but $\theta_{0a2}$ would not be. However, for most specifications for $q(t, u, w; \theta_{0a})$, the function $q(t, u, w; \theta_{0a})$ even restricted to $A_q$ uniquely determines $\theta_{0a}$ and thus (by extrapolation), the function the $q(t, u, w; \theta_{0a})$ on the complement of $A_q$ as well. It follows that, under such a specification, $q(t, u, w; \theta_{0a})$ is globally identified. We can make similar arguments for the global identification of $q^*(t, u, w; \theta_{0b})$ and $l(t, u, w; \theta_0)$. Identification of these functions will imply that $\theta_0$ is identified.

# 5 Causal Interpretation of $\eta_{0a}$ and $\theta_{0a}$

To talk about causality, it is very useful to think in terms of potential outcomes (Rubin, 1974). Define $Y^0(v)$ to be the logarithm of time to death if the subject, possibly contrary to fact, had been given treatment $v$ ($v = 0, 1$). Define $X^0(v)$ to be the logarithm of time to disease progression under treatment $v$. We will discuss causal interpretations when $X^0(v)$ is and is not well defined after $Y^0(v)$.

Under randomization, $V$ is independent of $(W, Y^0(0), Y^0(1), X^0(0) : X^0(0) \leq Y^0(0), X^0(1) : X^0(1) \leq Y^0(1))$.

## 5.1 Interpretation of $\eta_{0a}$

Under randomization, Model (1) can be written as

$$S_{Y^0(1)}(t|W = w) = S_{Y^0(0)}(r(t, w; \eta_{0a})|W = w) \tag{17}$$

where $S_{Y^0(z)}(\cdot|W = w)$ is the continuously, differential conditional survivor function of $Y^0(z)$ given $W = w$. As a result, $r(t, w; \eta_{0a})$ can be interpreted as the quantile-quantile mapping function between the distribution of death under treatment and the distribution of death

under no treatment, among subjects with covariates $w$ (see Figure 1). Since the quantile-quantile mapping function is a comparison of the distributions of potential outcomes on the same cohort of subjects, namely those with covariates $w$, the function and its parameters are causally interpretable. In contrast, the function $r^*(t, w; \eta_{0b})$ and $\eta_{0b}$ are not causally interpretable because $W$ is not randomized.

It is useful to note that

$$r(t, w; \eta_{0a}) = S^{-1}_{Y^0(0)}(S_{Y^0(1)}(t|W = w)|W = w)$$

and

$$r(Y^0(1), w; \eta_0) \overset{D(w)}{=} Y^0(0)$$

where $\overset{D(w)}{=}$ denotes equality in distribution given $W = w$.

Under the bivariate linear regression model (13,14), we have that $r(t, w; \eta_0) = t - \eta_{0a}$. So, we see that $\eta_{0a}$ is the constant additive shift in the death time distribuion due to the causal effect of treatment. Since $W$ was not randomized $\eta_{0b}$ does not have a casual interpretation.

## 5.2 Interpretation of $\theta_{0a}$

We now consider conditions under which $\theta_{0a}$ has a causal interpretation under randomization. In Section 5.2.1, we assume $X^0$ is well-defined for all subjects. In Section 5.2.2, we assume $X^0$ is defined only in subjects with $X^0 \leq Y^0$.

### 5.2.1 $X^0$ is well-defined for all subjects.

Under randomization, model (1,3) can be written as

$$S_{X^0(1)}(t|Y^0 = r^{-1}(u, w; \eta_{0a}), W = w) = S_{X^0(0)}(q(t, u, w; \theta_{0a})|Y^0 = u, W = w) \qquad (18)$$

Further, suppose $q(t,u,w;\theta_{0a}) = q(t,w;\theta_{0a})$ is free of $u$ for all $(t,u,w)$. Then,

$$S_{X^0(1)}(t|W=w) = S_{X^0(0)}(q(t,w;\theta_{0a})|W=w) \tag{19}$$

So, when $X^0$ is well defined for all subjects, $q(t,w;\theta_{0a})$ can be interpreted as the quantile-quantile mapping function between the distribution of progression under treatment and the distribution of progression under no treatment, among subjects with covariates $w$. That is, $q(t,w;\theta_{0a})$ is the causal conditional quantile-quantile function $S^{-1}_{X^0(0)}(S_{X^0(1)}(t|W=w)|W=w)$.

Now, suppose that $q(t,w,u;\theta_{0a})$ depends on $u$. Then, in general, $q(t,w,u;\theta_{0a})$ does not have a causal interpretation. This reflects the fact that, except under special circumstances described in the next subsection, the subset of the population defined by the event $\{Y^0 = r^{-1}(u,w;\eta_{0a}), W = w\}$ will generally differ in the distribution of counterfactuals $\{X^0(0), X^0(1)\}$ than the subset of the population defined by the event $Y^0 = u, W = w$. It follows that, even under the sharp null hypothesis

$$X^0(1) = X^0(0) \text{ with probability 1}$$

of no treatment effect of $X^0$, we would not expect $q(t,u,w;\theta_{0a}) = t$ to hold for all $(u,w)$. Thus, $q(t,u,w;\theta_{0a})$ would not have a causal interpretation.

To understand the implications of this result, consider the bivariate regression model specified in (11,12) where $h_1(w,u;\theta_{0a}) = h_1(w;\theta_{0a1})$ does not depend on $u$. Then, $q(t,u,w;\theta_{0a}) = t - h_1(w;\theta_{0a1})$. A goodness of fit test can be based on an expanded model in which $h_1(w,u;\theta_{0a}) = h_1(w;u;\theta_{0a1}) + \theta_{0a2}u$ in model (11,12). If the hypothesis $\theta_{0a2} = 0$ is rejected, then $q(t,u,w:\theta_{0a})$ in the expanded model does not have a causal interpretation. Even if the hypothesis is not rejected, and $q(t,u,w;\theta_{0a}) = t - h_1(w;\theta_{0a1})$ on $\mathcal{A}_q$, it is always

possible that

$$q(t,w,u;\theta_{0a}) = (t - h_1(w;\theta_{0a1})I((t,u,w) \in \mathcal{A}_q) + g(t,u,w;\theta_{0a3})I((t,u,w) \notin \mathcal{A}_q)$$

for some function $g(t,u,w;\theta_{0a3})$ defined on $\mathcal{A}_q^c$. If $g(t,u,w;\theta_{0a3}) \neq t - h_2(w;\theta_{0a1})$ for some $(t,u,w) \in \mathcal{A}_q^c$, then $q(t,u,w;\theta_{0a}) \neq t - h_1(w;\theta_{0a,1})$ for all $(t,u,w)$. Thus, $q(t,u,w;\theta_{0a})$ does not have a causal interpretation. Furthermore, since $q(t,u,w:\theta_{0a})$ is only identified on $\mathcal{A}_q$, we cannot test from the data whether $q(t,u,;:\theta_{0a}) \neq t - h_1(w;\theta_{0a1})$ on $\mathcal{A}_q^c$.

### 5.2.2 If $X^0$ is not defined for subjects with $X^0 \not\leq Y^0$

In the previous subsection we saw that if $X^0$ is well-defined for all subjects, then (i) $q(t,u,w;\theta_{0a})$ is well defined for all $(t,u,w)$ but only identified on $\mathcal{A}_q$ and (ii) if $q(t,u,w;\theta_{0a}) = q(t,w;\theta_{0a})$ for all $(t,u,w)$ then $q(t,w;\theta_0)$ had a causal interpretation as the causal quantile-quantile function $S^{-1}_{X^0(0)}(S_{X^0(1)|W=w}(t|W=w)$.

If $X^0$ is undefined for subjects with $X^0 \not\leq Y^0$, then (i) $q(t,u,w;\theta_{0a})$ is only defined and identified on $\mathcal{A}_q$ and (ii) the distributions $S_{X^0(1)}(t|W=w)$ and $S_{X^0(0)}(t|W=w)$ are not well-defined and thus, even if $q(t,u,w;\theta_{0a}) = q(t,w;\theta_{0a})$ where defined, $q(t,w;\theta_{0a})$ does not have a causal interpretation, except under special circumstances. These circumstances are that the following rank preservation assumption (Robins, 1995a) holds:

$$r(Y^0(1),w;\eta_0) = Y^0(0) \text{ with probability } 1(w)$$

This assumption implies that the rank ordering of subjects by $Y^0$ in the absence of treatment is preserved under treatment.

Under rank preservation and randomization, we can now write model (1,3) as

$$S_{X^0(1)}(t|Y^0(0)=u,W=w) = S_{X^0(0)}(q(t,u,w;\theta_{0a})|Y^0(0)=u,W=w) \qquad (20)$$

for $(t, u, w) \in \mathcal{A}_q$. Thus, $q(t, u, w; \theta_{0a})$ is, for $(t, u, w) \in \mathcal{A}_q$, the quantile-quantile mapping function between the distribution of disease progression under treatment and the distribution of disease progression under no treatment, among subjects who die at time $u$ under no treatment and have covariates $w$ (see Figure 2). Since the quantile-quantile mapping function is a comparison of the distributions of the potential outcomes on $X^0(1)$ and $X^0(0)$ for a fixed subset of subjects, namely those with die at time $u$ under no treatment and covariates $w$, the function and its parameters are causally interpretable. Thus, under rank preservation, $q(t, u, w; \theta_{0a})$ has a causal interpretation even when it depends on $u$. Robins (1986, Section 12.2, 1995ab, 2000) introduced this idea of defining causal contrasts conditional on functions of counterfactual survival times to deal with censoring by death. Frangakis and Rubin (2001) later discussed the same idea under the rubric of "principal stratification."

In many settings, the assumption of rank preservation is untenable. In that case, when $X^0$ is not defined for subjects with $X^0 \not\leq Y^0$, the causal effect of treatment on $X^0$ will generally not be identfied in any subset of the study population. Sensitivity analysis or bounding methods would be required,. As of now, the development of formal sensitivity analysis methods that use all available failure time information remains an open problem.

# 6 Estimation

## 6.1 More Notation

Our objective is to use the observed data to draw inference about $\beta_0 = (\theta_0, \eta_0)$. We assume $\beta_0$ lies in the interior of a compact set $\boldsymbol{\beta}$. For notational covenience, let

$$N_{Y^0}(u; \eta) = I(\tilde{M}(Y, Z; \eta) \leq u, \xi = 1)$$

and

$$N_{X^0}(u;\theta,\eta) = I(\tilde{L}(X,\tilde{M}(Y,Z;\eta),Z;\theta) \leq u, \delta = 1)$$

Let $H(\cdot,\cdot)$ denote the joint c.d.f. of $Z$ and $Y$, $F_{\tilde{M},\tilde{L}}(\cdot,\cdot)$ denote the joint c.d.f. of $\tilde{M}(\eta_0)$ and $\tilde{L}(\eta_0,\theta_0)$, and $\Lambda_{\tilde{M}}(\cdot)$ and $f_{\tilde{M}}(\cdot)$ denote the cumulative hazard and density functions of $\tilde{M}(\eta_0)$. Let $F_Z$ denote the cumulative distribution function of $Z$, and $S_C$ and $\Lambda_C$ denote the survivor and cumulative hazard functions of $C$, respectively. Let $N_C(t) = I(Y \leq t, \xi = 0)$ and $M_C(t) = N_C(t) - \int_{-\infty}^{t} I(Y \geq u) d\Lambda_C(u) du$. Throughout, true distributions are subscripted by zero and expectations without subscripts are taken with respect to the true law of the observed data, $P_0$.

## 6.2 Estimating Function for $\eta_0$

Consider the following estimating function for $\eta_0$;

$$\begin{aligned} U_1(O;\beta,H) &= \int_{-\infty}^{\infty} \{a(Z) - E_H[a(Z)|\tilde{M}(Y,Z;\eta) \geq u]\} dN_{Y^0}(u;\eta) \\ &= \int_{-\infty}^{\infty} \{a(Z) - \frac{E_H[a(Z) I(\tilde{M}(Y,Z;\eta) \geq u)]}{E_H[I(\tilde{M}(Y,Z;\eta) \geq u)]}\} dN_{Y^0}(u;\eta) \\ &= \int_{-\infty}^{\infty} \{a(Z) - \frac{\int_{z,y} a(z) I(\tilde{M}(y,z;\eta) \geq u) dH(z,y)}{\int_{z,y} I(\tilde{M}(y,z;\eta) \geq u) dH(z,y)}\} dN_{Y^0}(u;\eta) \end{aligned}$$

where $a(Z)$ is a bounded function of $Z$ with the same dimension as $\eta$.

Let $V_1(\beta,P) = \int U_1(o;\beta,H) dP(o)$, where $H$ is a function of $P$, the probability measure assigned to $O$. In the Appendix, we prove the following lemma;

**Lemma 2.** $V_1(\beta_0,P_0) = E[U_1(O;\beta_0,H_0)] = 0$.

The above estimating function is a generalization of one proposed by Tsiatis (1990) for classic linear regression with censored outcomes. For this special case, the estimating

function, with $a(Z) = Z$, reduces to:

$$U_1(O;\beta,H) = \int_{\infty}^{\infty} \left\{ Z - \frac{E_H[ZI(Y - Z'\eta \geq u)]}{E_H[I(Y - W'\eta \geq u)]} \right\} dN_{Y^0}(u)$$

Working the residual death time scale, this estimating function compares, for a subject who has been observed to die, his observed covariate vector to the expected value of the covariate vectors of subjects still at risk for death at that time. When covariates $W$ are excluded, the resulting estimating function is the same one proposed by Louis (1981), Wei and Gail (1983), and subsequently utilized by Lin, Robins, and Wei (1996) and Peng and Fine (2006).

## 6.3 Estimating Function of $\theta_0$

Assume $\eta_0$ is known. Now, consider the following estimating function:

$$\begin{aligned}
&U_2(O;\theta,\eta_0,H) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{ b(Z,s,u) - E_P[b(Z,s,u))|\tilde{L}(Y,\tilde{M}(Y,Z;\eta_0),Z;\theta) \geq u, \tilde{M}(Y,Z;\eta_0) = s, \xi = 1] \right\} dN_{X^0}(u;\theta,\eta_0) dN_{Y^0}(s;\eta_0) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{ b(Z,s,u) - \frac{E_P[b(Z,s,u)I(\tilde{L}(Y,s,Z;\theta) \geq u)|\tilde{M}(Y,Z;\eta_0) = s, \xi = 1]}{E_P[I(\tilde{L}(Y,s,Z;\theta) \geq u)|\tilde{M}(Y,Z;\eta_0) = s, \xi = 1]} \right\} dN_{X^0}(u;\theta,\eta_0) dN_{Y^0}(s;\eta_0) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{ b(Z,s,u) - \frac{E_P[b(Z,s,u)I(\tilde{L}(\tilde{M}^{-1}(s,Z;\eta_0),s,Z;\theta) \geq u)|\tilde{M}(Y,Z;\eta_0) = s, \xi = 1]}{E_P[I(\tilde{L}(\tilde{M}^{-1}(s,Z;\eta_0),s,Z;\theta) \geq u)|\tilde{M}(Y,Z;\eta_0) = s, \xi = 1]} \right\} dN_{X^0}(u;\theta,\eta_0) dN_{Y^0}(s;\eta_0) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{ b(Z,s,u) - \frac{E_H[b(Z,s,u)I(\tilde{L}(\tilde{M}^{-1}(s,Z;\eta_0),s,Z;\theta) \geq u)|\tilde{M}(Y,Z;\eta_0) \geq s]}{E_H[I(\tilde{L}(\tilde{M}^{-1}(s,Z;\eta_0),s,Z;\theta) \geq u)|\tilde{M}(Y,Z;\eta_0) \geq s]} \right\} dN_{X^0}(u;\theta,\eta_0) dN_{Y^0}(s;\eta_0) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{ b(Z,s,u) - \frac{E_H[b(Z,s,u)I(\tilde{L}(\tilde{M}^{-1}(s,Z;\eta_0),s,Z;\theta) \geq u, \tilde{M}(Y,Z;\eta_0) \geq s)]}{E_H[I(\tilde{L}(\tilde{M}^{-1}(s,Z;\eta_0),s,Z;\theta) \geq u, \tilde{M}(Y,Z;\eta_0) \geq s)]} \right\} dN_{X^0}(u;\theta,\eta_0) dN_{Y^0}(s;\eta_0) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{ b(Z,s,u) - \frac{\int_{z,y} b(z,s,u)I(\tilde{L}(\tilde{M}^{-1}(s,z;\eta_0),s,z;\theta) \geq u, \tilde{M}(y,z;\eta_0) \geq s) dH(z,y)}{\int_{z,y} I(\tilde{L}(\tilde{M}^{-1}(s,z;\eta_0),s,z;\theta) \geq u, \tilde{M}(y,z;\eta_0) \geq s) dH(z,y)} \right\} dN_{X^0}(u;\theta,\eta_0) dN_{Y^0}(s;\eta_0)
\end{aligned}$$

where $b(Z,s,u)$ is a specified function of $Z$, $s$ and $u$ with the same dimension as $\theta$.

Let $V_2(\beta,P) = \int U_2(o,\beta,H)dP(o)$. In the Appendix, we prove the following lemma:

**Lemma 3.** $V_2(\beta_0, P_0) = E[U_2(O; \beta_0, H_0)] = 0$.

In the context of the bivariate linear regression model, the above estimating function, with $b(Z, s, u) = Z$, reduces to

$$U_2(O; \theta, \eta, H) = \int_{-\infty}^{\infty} \int_{-\infty} \left\{ Z - \frac{E_H[Z I(Z'(\eta - \theta) \geq u - s, Y - Z'\eta \geq s)]}{E_H[I(Z'(\eta - \theta) \geq u - s, Y - Z'\eta \geq s)]} \right\} dN_{X^0}(u; \theta) dN_{Y^0}(s; \eta)$$

In this estimating function, one compares, for a subject who is observed to die and progress, his observed covariate vector to the expected value of the the covariate vectors for subjects who are observed to die and share the same observed death time (on the residual death time scale) and who at are risk for death at the progression time (on the residual progression scale).

Without covariates $W$, this latter estimating function is different than that proposed by Lin, Robins, and Wei (1996). The extension of their estimating function to covariates $W$ would be

$$\int_{-\infty}^{\infty} \{Z - E_P[Z | \tilde{X}(\beta) \geq u]\} d\tilde{N}_{X^0}(u; \beta)$$

where $\tilde{X}(\beta) = \min(X^0 - Z'\theta, Y - Z'\eta - d(\beta))$, $\tilde{\delta}(\beta) = I(X^0 - Z'\theta \leq Y - Z'\eta - d(\beta))$, $\tilde{N}_{X^0}(u; \beta) = I(\tilde{X}(\beta) \leq u, \tilde{\delta}(\beta) = 1)$, $d(\beta) = \sum_{k=1}^{q} (\theta_k - \eta_k)\{I(\theta_k - \eta_k > 0)u_k + I(\theta_k - \eta_k < 0)l_k\}$. The function $d(\beta)$ is chosen so that (1) it does not depend on $Z$ and (2) $\tilde{\delta}(\beta_0) = 1$ implies that $\delta = 1$. As a result, progression events are artificially censored. The above estimating function can be shown to have mean zero at the truth. The problem with their estimating function is that it is not practically useful for the setting in which high-dimensional covariates $W$ are included. In this setting, there will be excessive artificial censoring.

Arguing that excessive artificial censoring with high-dimensional covariates is caused by the fact that $d(\beta)$ is invariant across subjects, Peng and Fine (2006) introduced a U-statistic based estimating function that performs artificial censoring within pairs of subjects and then sums over all possible pairs. Their approach allows the trimming to be differ across

pairs of subjects, thus reducing the level of artificial censoring. Specifically, for pair $(i, j)$, the contribution to the estimating function is

$$(Z_i - Z_j)\left\{\tilde{\delta}_{ij}(\beta)I(\tilde{X}_j(\beta) \geq \tilde{X}_i(\beta) + \tilde{\delta}_{ji}(\beta)I(\tilde{X}_i(\beta) \geq \tilde{X}_j(\beta)\right\}$$

where $\tilde{X}_{ij}(\beta) = \min(X_i^0 - Z_i'\theta, Y_i - Z_i'\eta - d_{ij}(\beta))$, $\tilde{\delta}_{ij}(\beta) = I(X_i^0 - Z_i'\theta \leq Y_i - Z_i'\eta - d_{ij}(\beta))$ and $d_{ij}(\beta) = \max\{0, (\theta-\eta)'Z_i, (\theta-\eta)'Z_j\}$. The continuing need for artificial censoring could possibly adversely affect the efficiency of their estimator compared to ours. See below for further discussion of this issue.

### 6.4 Estimation of $\beta_0$

Let $V(\beta, P) = (V_1(\beta, P), V_2(\beta, P))'$. Since $V(\beta, P_0)$ is continuous and differentiable in $\beta$ and $V(\beta_0, P_0) = 0$, it is natural to think of estimating $\beta_0$ as the solution to $V(\beta, P_n) = 0$, where $P_n$ is the empirical distribution of the observed data. Unfortunately, $V(\beta, P_n)$ is discontinuous in $\beta$. Thus, we propose to estimate $\beta_0$ as the maximizer, $\beta_n$, of $Q_n(\beta) = -V(\beta, P_n)'V(\beta, P_n)$. In the next section, we show that $\beta_n$ is a consistent and asymptotically normal estimator of $\beta_0$ and we construct a consistent estimator of its asymptotic variance.

## 7 Large Sample Theory

Our results rely on high-level framework for large sample estimation developed by Newey and McFadden (NM;1994). For convenience, in Section 5.1, we reproduce their Theorems 7.1, 7.2 (with $W = I$), and 7.4. In Section 5.2, we introduce regularity conditions. In section 5.3, we present two lemmas that pave our way for the main large sample results, which are stated as two Theorems in Section 5.4. In what follows, let $Q_0(\beta) = -V(\beta, P_0)'V(\beta, P_0)$.

## 7.1 Theorems of Newey and McFadden

**Theorem 1 (7.1 of NM).** *If there is a function $Q_0(\beta)$ such that (i) $Q_0(\beta)$ is uniquely maximized at $\beta_0$; (ii) $\boldsymbol{\beta}$ is compact; (iii) $Q_0(\beta)$ is continuous; (iv) $Q_n(\beta)$ converges uniformly to $Q_0(\beta)$, then $\beta_n$ converges in probability to $\beta_0$*

**Theorem 2 (7.2 of NM).** *Suppose (i) $\beta_n$ converges in probability to $\beta_0$; (ii) $V(\beta_0, P_0) = 0$; (iii) $V(\beta, P_0)$ is differentiable at $\beta_0$ with derivative $G(\beta_0, P_0)$; (iv) $\beta_0$ is in the interior of $\boldsymbol{\beta}$; (v) $\sqrt{n}V(\beta_0, P_n) \xrightarrow{D} N(0, \Sigma(\beta_0, P_0))$; ( vi) for any $\epsilon_n \to 0$, $\sup_{\|\beta-\beta_0\|\leq\epsilon_n} \sqrt{n}\|V(\beta, P_n) - V(\beta_0, P_n) - V(\beta, P_0)\|/\{1 + \sqrt{n}\|\beta - \beta_0\|\} \xrightarrow{P} 0$. Then,*

$$\sqrt{n}(\beta_n - \beta_0) \xrightarrow{D} N(0, G(\beta_0, P_0)^{-1}\Sigma(\beta_0, P_0)G(\beta_0, P_0)^{-1'}).$$

Let $G^\dagger(\beta_n, P_n))$ be a numerical derivative estimator of $G(\beta_0, P_0)$, where the $j$th column of $G^\dagger(\beta, P_n))$ is

$$G_j^\dagger(\beta, P_n) = \{V(\beta + e_j\epsilon_n, P_n) - V(\beta - e_j\epsilon_n, P_n)\}/(2\epsilon_n) \tag{21}$$

and $e_j$ is the $j$th unit vector.

**Theorem 3 (7.4 of NM).** *Suppose $\epsilon_n \to 0$ and $\epsilon_n\sqrt{n} \to \infty$. If the conditions of Theorems 1 and 2 hold, then $G^\dagger(\beta_n, P_n) \xrightarrow{P} G(\beta_0, P_0)$.*

## 7.2 Regularity Conditions

Let

$$f_1(z, y, u; \beta) \equiv I(\tilde{M}(y, z; \eta) \geq u)$$
$$f_2(z, y, u, s; \beta) \equiv I(\tilde{L}(\tilde{M}^{-1}(s, z; \eta), s, z; \theta) \geq u, \tilde{M}(y, z; \eta) \geq s).$$

Partition $Z$ into its continuous components $Z_c$ and its discrete components $Z_d$.

RC1: Let the parameter space of $\beta$ be a Euclidean sphere, that is, $\boldsymbol{\beta} = \{\beta : \| \beta \| \leq M\}$ for some $M < \infty$.

RC2: $\beta_0$ lies in the interior of $\boldsymbol{\beta}$

RC3: $V(\beta, P_0) = 0$ has only one solution in $\boldsymbol{\beta}$.

RC4: For some $L$, $\|Z\| < L$ with probability 1.

RC5: Assume that

$$\mathcal{G}_1 \equiv \{O \to U_1(O, \beta, H) : \beta, H\} \text{ is a uniformly bounded } P_0\text{-Donsker class,}$$

$$\mathcal{F}_1 \equiv \{(z, y) \to f_1(z, y, u; \beta) : u, \beta\} \text{ is a uniformly bounded } H_0\text{-Donsker class,}$$

and with probability 1,

$$I(\xi = 1) \int f_1(z, y, \tilde{M}(Y, Z; \eta); \beta) dH_0(y, z) > \delta > 0$$

for some $\delta > 0$ uniformly in $\beta \in \boldsymbol{\beta}$.

RC6: Assume that

$$\mathcal{G}_2 \equiv \{O \to U_2(O; \beta, H) : \beta, H\} \text{ is a uniformly bounded } P_0\text{-Donsker class,}$$

$$\mathcal{F}_2 \equiv \{(z, y) \to f_2(z, y, u, s; \beta) : u, s, \beta\} \text{ is a uniformly bounded } H_0\text{-Donsker class,}$$

$$\mathcal{F}_b \equiv \{z \to b(z, s) : s\} \text{ is a uniformly bounded } H_0\text{-Donsker class,}$$

and with probability 1,

$$I(\delta = \xi = 1) \int f_2(z, y, \tilde{L}(X, \tilde{M}(Y, Z; \eta), Z; \theta), \tilde{M}(Y, Z; \eta); \beta) dH_0(y, z) > \delta$$

for some $\delta > 0$, uniformly in $\beta \in \boldsymbol{\beta}$.

RC7: Within all the possible levels of $(\delta, \xi, Z_d)$, the sub-distribution of $(X, Y, Z_c)$ is absolutely continuous w.r.t. Lebesgue measure.

RC8: The derivative (with respect to $\beta$) of $V(\beta, P_0)$, $G(\beta, P_0)$, exists, is invertible and continuous at $\beta_0$

The regularity conditions RC1-RC4 are standard, and state, beyond boundedness conditions on the parameter space for $\beta$ and Z, that the estimating equation asymptotically uniquely identifies the true parameter value. RC7 and RC8 are also standard smoothness condition necessary to make our estimator a smooth enough function of the empirical distribution so that it is asymptotically linear, and thus asymptotically normally distributed (that is, it behaves as a sample mean in first order).

The Donsker class conditions RC5 and RC6 are necessary in order to establish the wished asymptotic linearity. For a given class (e.g., $\mathcal{G}_1$), it holds, if each of the functions in this class, considered as a multivariate real valued function in $(x, y, z_c)$, given $(\delta, \xi, Z_d)$, have a uniform sectional variation norm bounded by a universal constant (Gill, van der Laan, Wellner, 1994). The uniform sectional variation norm of a multivariate function is defined as the supremum over all its sections of the regular variation norm of the given section of the function, where a section of a multivariate function is one of the lower dimensional functions one obtains by fixing one or more of the coordinates. To establish such a result one typically assumes that, within levels of $(\delta, \xi, Z_d)$, the support of $(X, Y, Z_c)$ is compact. In words, these classes are Donsker classes if, within levels of $(\delta, \xi, Z_d)$, they are reasonably smooth (regarding changes in value, e.g. number of jumps) functions in $(X, Y, Z_c)$, and that $(X, Y, Z_c)$ has compact support. We refer to van der Vaart, Wellner (1996) for many more examples of Donsker classes.

## 7.3 Two Lemmas

The following lemmas provide us with a linear expansion of $V(\beta, P_n) - V(\beta, P_0)$ uniform in $\beta$, which will be fundamental in proving consistency and asymptotic normality of $\beta_n$. The proof of Lemma 4 is provided in the Appendix. Lemma 3 follows the same logic. We will use the notation $Pf = \int f(x)dP(x)$.

**Lemma 4.** *Under regularity conditions RC4 and RC5,*

$$V_1(\beta, P_n) - V_1(\beta, P_0) = (P_n - P_0)IC_1(\cdot \mid \beta, P_0) + R_{n,1}(\beta, P_0),$$

*where*

$$IC_1(O \mid \beta, P_0) = U_1(O; \beta, H_0) - \int \frac{a(Z)f_1(Z, Y, u; \beta)}{\int f(z, y, u; \beta)dH_0(y, z)} E[dN_{Y^0}(u, \eta_0)] + \int f_1(Z, Y, u; \beta) \frac{\int a(z)f_1(z, y, u; \beta)dH_0(y, z)}{\left(\int f_1(z, y, u; \beta)dH_0(y, z)\right)^2} E[dN_{Y^0}(u, \eta_0)],$$

$\{O \rightarrow IC_1(O \mid \beta, P_0) : \beta \in \boldsymbol{\beta}\}$ *is a* $P_0$*-Donsker class, and* $\sup_{\beta \in \boldsymbol{\beta}} \parallel R_{n,1}(\beta, P_0) \parallel = o_{P_0}(1/\sqrt{n})$.

**Lemma 5.** *Under regularity conditions RC4 and RC6,*

$$V_2(\beta, P_n) - V_2(\beta, P_0) = (P_n - P_0)IC_2(\cdot \mid \beta, P_0) + R_{n,2}(\beta, P_0),$$

*where*

$$IC_2(O \mid \beta, P_0) = U_2(O; \beta, H_0) - \int \frac{b(Z, s, u)f_2(Z, Y, u, s; \beta)}{\int f_2(z, y, u, s; \beta)dH_0(y, z)} E[(dN_{X^0}(u, \theta)dN_{Y^0}(s, \eta_0))] + \int f_2(Z, Y, u, s; \beta) \frac{\int b(z, s)f_2(z, y, u, s; \beta)dH_0(y, z)}{\left(\int f_2(z, y, u, s; \beta)dH_0(y, z)\right)^2} E[(dN_{X^0}(u, \theta)dN_{Y^0}(s, \eta_0))],$$

$\{O \rightarrow IC_2(O \mid \beta, P_0) : \beta \in \boldsymbol{\beta}\}$ *is a* $P_0$*-Donsker class, and* $\sup_{\beta \in \boldsymbol{\beta}} \parallel R_{n,2}(\beta) \parallel = o_{P_0}(1/\sqrt{n})$.

Thus, we have that

$$V(\beta, P_n) - V(\beta, P_0) = (P_n - P_0)IC(\cdot \mid \beta, P_0) + R_n(\beta, P_0),$$

where $IC(\cdot \mid \beta, P_0) = (IC_1(\cdot \mid \beta, P_0), IC(\cdot \mid \beta, P_0))'$ and $R_n(\beta, P_0) = (R_{n,1}(\beta, P_0), R_{n,2}(\beta, P_0))'$.

## 7.4 Large Sample Results

### 7.4.1 Consistency

Under conditions RC1 and RC3, we know that conditions (i) and (ii) of Theorem 1 are satisfied. Furthermore, under condition RC7, we have that $V(\beta, P_0)$ is continuous in $\beta$, and thus condition (iii) of the general consistency Theorem 1 holds. Finally, under RC4-RC6 (i.e, the conditions of Lemmas 3 and 4, we have $\sup_{\beta \in \boldsymbol{\beta}} \parallel V(\beta, P_n) - V(\beta, P_0) \parallel = O_{P_0}(1/\sqrt{n})$. This tells us that condition (iv) of Theorem 1 holds. Thus, we are in the position to state the following theorem:

**Lemma 6.** *Under conditions RC1,RC3-RC7, Theorem 1 applies and $\beta_n$ is asymptotically consistent.*

### 7.4.2 Asymptotic Normality

We now work on the verification of the conditions of the general asymptotic normality Theorem 2. Lemma 5 verifies condition (i) of Theorem 2. Lemmas 1 and 2 verify condition (ii). Conditions (iii) and (iv) hold by assumptions RC2 and RC8. From Lemmas 3 and 4, we know that

$$\sqrt{n}\{V(\beta_0, P_n) - V(\beta_0, P_0)\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IC(O_i \mid \beta_0, P_0) + o_{P_0}(1),$$

where $IC(O \mid \beta, P_0) = (IC_1(O \mid \beta, P_0), IC_2(O \mid \beta, P_0))^\top$. This proves condition (v) of Theorem 2 with $\Sigma(\beta_0, P_0) = E[IC(O \mid \beta_0, P_0)IC(O \mid \beta_0, P_0)^\top]$.

Regarding condition (vi) of Theorem 2, we note that

$$\begin{aligned}\delta(n) &\equiv V(\beta_0 + \epsilon_n, P_n) - V(\beta_0, P_n) - V(\beta_0 + \epsilon_n, P_0) \\ &= V(\beta_0 + \epsilon_n, P_n) - V(\beta_0 + \epsilon_n, P_0) - \{V(\beta_0, P_n) - V(\beta_0, P_0)\}.\end{aligned}$$

This shows that the term $\delta(n)$ in condition (vi) can be expected to be second order for any sequence $\epsilon_n$ converging to zero, and thus that condition (vi) of Theorem 2 is a natural condition. By application of the linear expansions of Lemmas 3 and 4 at $\beta_0 + \epsilon_n$ and $\beta_0$, respectively, we have that

$$\begin{aligned}V(\beta_0 + \epsilon_n, P_n) - V(\beta_0 + \epsilon_n, P_0) &= (P_n - P_0)IC(\cdot \mid \beta_0 + \epsilon_n, P_0) + o_{P_0}(1/\sqrt{n}) \\ V(\beta_0, P_n) - V(\beta_0, P_0) &= (P_n - P_0)IC(\cdot \mid \beta_0, P_0) + o_{P_0}(1/\sqrt{n}).\end{aligned}$$

Thus,

$$\delta(n) = (P_n - P_0)\{IC(\cdot \mid \beta_0 + \epsilon_n, P_0) - IC(\cdot \mid \beta_0, P_0)\} + o_{P_0}(1/\sqrt{n}).$$

Lemmas (3) and (4) also teach us that $\{O \to IC(O \mid \beta, P_0) : \beta\}$ is a $P_0$-Donsker class. It is well known from empirical process theory that, if 1) $f_n$ falls in a Donsker class with probability tending to 1, and 2) $\int f_n^2(o)dP_0(o)$ converges to zero in probability, then $\int f_n(o)d(P_n - P_0)(o) = o_{P_0}(1/\sqrt{n})$ (van der Vaart, Wellner, 1996). Consequently, $\delta(n) = o_{P_0}(1/\sqrt{n})$ if $\int\{IC(O \mid \beta_0 + \epsilon_n, P_0) - IC(O \mid \beta_0, P_0)\}^2 dP_0(O) \to 0$ in probability. It is easy to verify that the latter holds under the regularity conditions (in particular, RC8). Now, we can state the following lemma:

**Lemma 7.** *Under conditions RC1-RC8, Theorem 2 applies and $\beta_n$ is asymptotically linear with influence curve*

$$-G(\beta_0, P_0)^{-1}IC(O \mid \beta_0, P_0).$$

*In particular,* $\sqrt{n}(\beta_n - \beta_0) \stackrel{D}{\Rightarrow} N(O, G(\beta_0, P_0)^{-1}\Sigma G(\beta_0, P_0)^{-1'})$*, where* $\Sigma(\beta_0, P_0) = E[IC(O \mid \beta_0, P_0)IC(O \mid \beta_0, P_0)']$.

### 7.4.3 Variance Estimation

Under RC1-RC8, Theorem 3 tells us that we can estimate $G(\beta_0, P_0)$ by $G^\dagger(\beta_n, P_n)$ given in Equation (3). We can also estimate $\Sigma(\beta_0, P_0)$ by $\Sigma(\beta_n, P_n)$. Thus, we can consistently estimate the asymptotic variance of $\beta_n$ by $G^\dagger(\beta_n, P_n)^{-1}\Sigma(\beta_n, P_n)G^\dagger(\beta_n, P_n)^{-1\top}$. The consistency of this estimate of the asymptotic variance follows under the same conditions as we needed for the asymptotic linearity of $\beta_n$.

## 8 Simulation Studies

We conducted two simulation studies to evaluate the performace of our estimator. For each study, we simulated three covariates ($Z = (Z^{(1)}, Z^{(2)}, Z^{(3)})'$) in a sequential fashion: $Z^{(1)}$ was generated as a Bernoulli random variable with probability 0.5, $Z^{(2)}$ given $Z^{(1)}$ was generated as a truncated (at 2 and 7.5) normal random variable with mean $4.5 + 0.5Z_1$ and variance 1, and $Z^{(3)}$ given $Z^{(1)}$ and $Z^{(2)}$ was generated as a Bernoulli random variable with probability $0.3 + 0.2 * Z_1$. The true value of $\theta_0$ and $\eta_0$ in (1) and (2) was set equal to $(0.4, -0.1, -0.2)$ and $(0.2, -0.3, -0.4)$, respectively. We assumed that the censoring time was independent and followed the log of a uniform (0,c) random variable. 1000 simulated datasets of 250 independent subjects were created for each study.

In the first simulation study, $\exp(\epsilon)$ and $\exp(\nu)$ were generated from a bivariate exponential with hazard rates 1.0 and 0.2, respectively, and correlation coefficient 0.25. We set $c =$??. On average, the censoring rate for death and disease progression was 23.2% and 44.36%, respectively; both progression and death were observed on 40.4% of subjects. 1.8% of the simulations failed to converge. The results of the converged simulations are presented

to Table 2 (a). As we see, our estimator of the regression parameters has low bias. This is seen by comparing the Monte Carlo average (M.C. Avg.) over simulations to the truth. In addition, our influence-based standard error estimator works well; the Monte Carlo average of the estimated standard errors (Avg. Est. S.E.) is close to the Monte Carlo standard deviation (M.C. S.D.) of the simulated estimates. Finally, the coverage rate (Cov. Prob.) of 95% Wald-based confidence is close to the nominal level.

In the second simulation study, $\epsilon$ and $\nu$ were generated from a bivariate normal with means 0.0 and 1.5, respectively, variances 1 and correlation 0.25. Here, we set $c = 8$. On average, the censoring rate for death and disease progression was 21.6% and 44.1%, respectively; both progression and death were observed on 44.8% of subjects. 1.1% of the simulations failed to converge. As seen in Table 2 (b), our estimation procedure performs well.

# 9 Analysis of Glioma Trial

In our analysis of the glioma trial, $V$ is the indicator of assignment to the carmustine ploymer arm and $W$ is a $10 \times 1$ vector of prognostic factors including resection greater than 75%, age in years (continuous), white, Karnofsky score greater than 70, local radiation, previous use of nitrosoureas, active tumor histology, anaplastic astrocytoma subtype, oliodendroglioma subtype, and other subtype. To evaluate whether a causal interpretation can be (statistically) ruled out, we first fit model (11,12) with $g_1(W;\eta_{0a}) = \eta_{0a}$, $g_2(W;\eta_{0b}) = W'\eta_{0b}$, $h_1(W,u;\theta_{0a}) = \theta_{0a1} + \theta_{0a2}u$ and $h_2(W,u;\theta_{0b} = W'\theta_{0b}$ and tested whether the null hypothesis $\theta_{0a2} = 0$ can be rejected at the 0.05 level. In our estimating function, we let $a(Z) = Z$ and $b(Z,s,u) = (Z', Vu)'$. The resulting estimate of $\theta_{0a2}$ is -0.0079, with standard error 0.21. Using a Wald test, we are unable to reject the null.



Next, we fit the reduced model (13,14). For this model, we used our estimating

function approach (SRL) and that of Peng and Fine (2006;PF). The results of are shown in Table 2. Both approaches produce the same inference for regression coefficients of the death model (see final column). Carmustine ploymer appears to increase the time to death relative to placebo polymer (95% CI: 0.06,0.50). These results are consistent with Brem *et al.* (1995), who fit a proportional hazards regression version of our model for death and found an adjusted hazard ratio of 0.67 (95% CI: 0.51 -0.90).

For cerebellar progression, inference about the effect of carmustine ploymer depends on the estimation technique. Our estimating function approach, using $a(Z) = Z$ and $b(Z, s, u) = Z$, shows, under rank preservation, that carmustine polymer also increases the time to progression (95%: 0.07,0.68). In contrast, the approach of Peng and Fine (2006), who use a resampling-based standard error estimator, do not suggest a benefit of treatment with carmustine on cerebellar progression (95% -0.23,0.92) [1]. In fact , 9 of 11 standard error estimates are smaller using our approach than that of Peng and Fine (2006). Our progression results are consistent with Westphal *et al.* (2003), who reported the results of a follow-up clinical trial.

# 10 Discussion

In this paper we have imbedded the bivariate linear regression model for the analysis semi-competing risks data into a more general quantile-quantile regression model. The use of this more general model allowed us to clarify the conditions under which the model parameters have a causal interprestation. Specifically, in settings in which $X^0$ is well-defined even when $X^0 \not< Y^0$, we showed in Section 5.2.1 that, even in a randomized experiment, if the quantile-quantile function $q(t, u, w; \theta_{0a})$ depends on $u$, the parameter $\theta_{0a}$ does not have a causal interpretation and, in fact, no progression-related causal contrast is identified from semi-competing risks data. Now the bivariate linear regression model is a special case of

[1] We are grateful to Limin Peng for conducting this analysis

our quantile-quantile model for which $q(t,u,w;\theta_{0a})$ is free of $u$ and thus $\theta_{0a}$ has a causal interpretation. However, the dependence of $q(t,u,w;\theta_{0a})$ on $u$ is testable from the data. Thus we recommend that one never use a bivariate linear regression model without first testing whether $q(t,u,w;\theta_{0a})$ is free of $u$. If the test rejects, one should also reject the bivariate linear regression model and any causal interpretation for $\theta_{0a}$.

In settings in which $X^0$ is not defined when $X^0 \not< Y^0$, we showed that, in a randomized study, the quantile-quantile function $q(t,u,w;\theta_{0a})$ has a causal interpretation, even if it depends on $u$, under a strong rank preservation assumption. As this rank preservation assumption is often unlikely to hold, it would be useful to develop a sensitivity analysis methodology for our model depending on a non-identified sensitivity-parameter that could be used estimate the quantile -quantile function linking the distribution of the progression counterfactual $X^0(1)$ with the counterfactual $X^0(0)$ given $Y^0(0) = u, W = w$ on the set where $X^0(1)$ and $X^0(0)$ are well-defined.

Finally in this paper we have not attempted to construct a semiparametric efficient estimator of $\theta_{0a}$, although we plan to do so in subsequent work. Even so, in our empirical example, our estimator of $\theta_{0a}$ had an estimated variance much smaller than did the rank estimator of Peng and Fine (2006). The reader may find this suprising because Peng and Fine (2006) reported results of a simulation study that, in a setting of 2-dimensional $Z$, demonstated that their rank estimator had better finite-sample effiiciency properties than even the semiparametric efficient estimator of Tsiatis in the smaller model that assumes that $X^0$ and $Y^0$ are independent given $Z$ . However, the semiparametric efficient estimator depends on a smoothed estimate of the derivative of the log hazard of $X^0 - \theta_0^T Z$. Peng and Fine's (2006) method of selecting their smoothing parameter resulted in a small bandwidth and thus large second order variance terms in the expansion of the semiparametric efficient estimator, resulting in a loss of efficiency in finite samples. The performance of the semiparametric efficient estimator could have been improved, possibly considerably, by either using a larger bandwidth or by using a low dimensional parametric model for the hazard

of $X^0 - \theta_0^T Z$ that included the truth. This latter estimator would be referred to as locally, but not globally, semiparamtric efficient. Since we did not attempt to construct a semiparametric efficient estimator, our estimator did not require us to estimate any non- root-n estimable function (such as the derivative of a log hazard) and thus does not include second order variance terms in its expansion. As a consequence, it would not be suprising for our estimator to be more efficient than Peng and Fine's (2006), although we have not carried out either a detailed comparison, either analytically or by simulation.

## Appendix

**Proof of Lemma 1:**

Part (a) follows since

$$\begin{aligned} P[\tilde{R}(\eta_{0a}) > t | V=1, W=w] &\overset{(7)}{=} P[r(Y^0, W; \eta_{0a}) > t | V=1, W=w] \\ &= P[Y^0 > r^{-1}(t, w; \eta_{0a}) | V=1, W=w] \\ &\overset{(1)}{=} S_{Y^0}(t | V=0, W=w) \\ P[\tilde{R}(\eta_{0a}) > t | V=0, W=w] &\overset{(7)}{=} S_{Y^0}(t | V=0, W=w) \end{aligned}$$

Part (b) follows since

$$P[\tilde{M}(\eta_0) > t | V=1, W=w] \overset{(8)}{=} P[Y^0 > m^{-1}(t, w; \eta_0) | V=1, W=w] \overset{(5)}{=} S_{Y^0}(t | V=0, W=0)$$

$$P[\tilde{M}(\eta_0) > t | V=0, W=w] \overset{(8)}{=} P[Y^0 > r^{*-1}(t, w; \eta_{0b}) | V=0, W=w] \overset{(2)}{=} S_{Y^0}(t | V=0, W=0)$$

Part (c) follows since

$$
\begin{aligned}
&P[\tilde{Q}(\eta_{0a},\theta_{0a}) > t | V=1, \tilde{R}(\eta_{0a}) = u, W = w] \\
&\qquad \overset{(7,9)}{=} P[q(X^0, \tilde{R}(\eta_{0a}), w; \theta_{0a}) > t | V=1, r(Y^0, w; \eta_{0a}) = u, W = w] \\
&\qquad = P[X^0 > q^{-1}(t, \tilde{R}(\eta_{0a}), w; \theta_{0a}) | V=1, Y^0 = r^{-1}(u, w; \eta_{0a}), W = w] \\
&\qquad \overset{(3)}{=} S_{X^0}(t | V=0, Y^0 = u, W = w) \\
&P[\tilde{Q}(\eta_{0a},\theta_{0a}) > t | V=0, \tilde{R}(\eta_{0a}) = u, W = w] \overset{(7,9)}{=} S_{X^0}(t | V=0, Y^0 = u, W = w)
\end{aligned}
$$

Part (d) follows since

$$
\begin{aligned}
&P[\tilde{L}(\eta_0,\theta_0) > t | V=1, \tilde{M}(\eta_0) = u, W = w] \\
&\qquad \overset{(8)}{=} P[\tilde{L}(\eta_0,\theta_0) > t | V=1, Y^0 = m^{-1}(u, w; \eta_0), W = w] \\
&\qquad \overset{(10)}{=} P[X^0 > l^{-1}(t, u, w; \eta_{0b}, \theta_0) | V=1, Y^0 = m^{-1}(u, w; \eta_0), W = w] \\
&\qquad \overset{(6)}{=} S_{X^0}(t | V=0, Y^0 = u, W = 0) \\
&P[\tilde{L}(\eta_0,\theta_0) > t | V=0, \tilde{M}(\eta_0) = u, W = w] \\
&\qquad \overset{(8,10)}{=} P[q^*(X^0, u, w; \theta_{0b}) > t | V=0, Y^0 = r^{*-1}(u, w; \eta_{0b}), W = w] \\
&\qquad = P[X^0 > q^{*-1}(t, u, w; \theta_{0b}) | V=0, Y^0 = r^{*-1}(u, w; \eta_{0b}), W = w] \\
&\qquad \overset{(4)}{=} S_{X^0}(t | V=0, Y^0 = u, W = 0)
\end{aligned}
$$

Part (e) follows from results (b) and (d).

**Proof of Lemma 2:**

$$
\begin{aligned}
V_1(\eta_0, P_0) &= E[\int_{-\infty}^{\infty}\{a(Z) - \frac{E[a(Z)I(\tilde{M}(Y,Z;\eta_0) \geq u)]}{E[I(\tilde{M}(Y,Z;\eta_0) \geq u)]}\}dN_{Y^0}(u;\eta_0)] \\
&= \int_{-\infty}^{\infty}\{E[a(Z)dN_{Y^0}(u;\eta_0)] - \frac{E[a(Z)I(\tilde{M}(Y,Z;\eta_0) \geq u)]}{E[I(\tilde{M}(Y,Z;\eta_0) \geq u)]}E[dN_{Y^0}(u;\eta_0)]\} \\
&= \int_{-\infty}^{\infty}\{E[a(Z)I(\tilde{M}(Y,Z;\eta_0) \geq u)]d\Lambda_{\tilde{M}}(u) - \\
&\qquad \frac{E[a(Z)I(\tilde{M}(Y,Z;\eta_0) \geq u)]}{E[I(\tilde{M}(Y,Z;\eta_0) \geq u)]}E[I(\tilde{M}(Y,Z;\eta_0) \geq u]d\Lambda_{\tilde{M}}(u)\} \\
&= 0
\end{aligned}
$$

**Proof of Lemma 3:**

$$
\begin{aligned}
&V_2(\theta_0, \eta_0, P_0) \\
&= E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{b(Z,s,u) - E[b(Z,s,u)|\tilde{L}(Y,\tilde{M}(Y,Z;\eta_0),Z;\eta_0) \geq u, \tilde{M}(Y,Z;\eta_0) = s, \xi = 1]\}\right. \\
&\qquad\qquad \left. dN_{X^0}(u;\theta_0)dN_{Y^0}(s;\eta_0)\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E[b(Z,s,u)dN_{X^0}(u;\theta_0)dN_{Y^0}(s;\eta_0)] - \qquad (22) \\
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{E[b(Z,s,u)I(\tilde{L}(Y,\tilde{M}(Y,Z;\eta_0),Z;\eta_0) \geq u, \tilde{M}(Y,Z;\eta_0) = s, \xi = 1)]}{E[I(\tilde{L}(Y,\tilde{M}(Y,Z;\eta_0),Z;\eta_0) \geq u, \tilde{M}(Y,Z;\eta_0) = s, \xi = 1)]}E[dN_{X^0}(u;\theta_0)dN_{Y^0}(s;\eta_0)] \qquad (23)
\end{aligned}
$$

Now, note that, for any function $k(Z,s)$,

$$
\begin{aligned}
&E[k(Z,s)dN_{X^0}(u;\theta_0)dN_{Y^0}(s;\eta_0)] \\
&= E[I(\tilde{M}(Y_0,Z;\eta_0) \in ds)I(\tilde{L}(X^0,\tilde{M}(Y^0,Z;\eta_0),Z;\theta_0) \in ds)k(Z,s)I(\tilde{M}^{-1}(s,Z;\eta_0) \geq \tilde{L}^{-1}(u,s,Z;\theta_0), C \geq \tilde{M}^{-1}(s,Z;\eta_0))] \\
&= F_{\tilde{M},\tilde{L}}(du,ds)E[k(Z,s)I(\tilde{M}^{-1}(s,Z;\eta_0) \geq \tilde{L}^{-1}(u,s,Z;\theta_0), C \geq \tilde{M}^{-1}(s,Z;\eta_0))]
\end{aligned}
$$

and

$$
\begin{aligned}
&E[k(Z,s)I(\tilde{L}(Y,\tilde{M}(Y,Z;\eta_0),Z;\eta_0)\geq u,\tilde{M}(Y,Z;\eta_0)=s,\xi=1)] \qquad (24)\\
&=E[k(Z,s)I(\tilde{M}(Y^0,Z;\eta_0)\in ds)I(\tilde{M}^{-1}(s,Z;\eta_0)\geq\tilde{L}^{-1}(u,s,Z;\theta_0),C\geq\tilde{M}^{-1}(s,Z;\eta_0))]\\
&=F_{\tilde{M}}(ds)E[k(Z,s)I(\tilde{M}^{-1}(s,Z;\eta_0)\geq\tilde{L}^{-1}(u,s,Z;\theta_0),C\geq\tilde{M}^{-1}(s,Z;\eta_0))]
\end{aligned}
$$

where the last equalities in (6) and (7) follow by the joint independence of $(\tilde{M}(\eta_0),\tilde{L}(\eta_0,\theta_0))$ and $(C,Z)$. Using (6) and (7) with $k(Z,s)$ set equal to $b(Z,s,u)$ and 1, respectively, we can then plug the relevant expectations into (4) and (5). We then obtain that

$$
\begin{aligned}
V_2(\theta_0,\eta_0,P_0)=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E[b(Z,s,u)I(\tilde{M}^{-1}(s,Z;\eta_0)\geq\tilde{L}^{-1}(u,s,Z;\theta_0),C\geq\tilde{M}^{-1}(s,Z;\eta_0))]F_{\epsilon,\nu}(du,ds)-\\
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{E[b(Z,s,u)I(\tilde{M}^{-1}(s,Z;\eta_0)\geq\tilde{L}^{-1}(u,s,Z;\theta_0),C\geq\tilde{M}^{-1}(s,Z;\eta_0))]}{E[I(\tilde{M}^{-1}(s,Z;\eta_0)\geq\tilde{L}^{-1}(u,s,Z;\theta_0),C\geq\tilde{M}^{-1}(s,Z;\eta_0))]}\times\\
&E[I(\tilde{M}^{-1}(s,Z;\eta_0)\geq\tilde{L}^{-1}(u,s,Z;\theta_0),C\geq\tilde{M}^{-1}(s,Z;\eta_0))]F_{\epsilon,\nu}(du,ds)
\end{aligned}
$$

Note the cancellation of the cancelation of like terms in the numerator and the denominator in the second term of the substraction. The remaining terms in the subtraction are now identical, yielding that $V_2(\theta_0,\eta_0,P_0)=0$.

**Proof of Lemma 5:**

Note that we can write $E_{P_n}[U_2(O;\beta,H_n)]-E[U_2(O;\beta,H_0)]$ is equal to

$$
E_{P_n-P_0}[U_2(O;\beta,H_n)-U_2(O;\beta,H_0)]+E_{P_n-P_0}[U_2(O;\beta,H_0)]+E[U_2(O;\beta,H_n)-U_2(O;\beta,H_0)]. \qquad (25)
$$

Under the condition that $\mathcal{G}_2$ is a $P_0$-Donsker class (RC5), we know by empirical process

theory (van der Vaart and Wellner, 1996), that if

$$\int \{U_2(O;\beta,H_n) - U_2(O;\beta,H_0)\}^2 dP_0(O) \xrightarrow{P_0} 0 \tag{26}$$

then the first and second terms in (25) are $o_{P_0}(1/\sqrt{n})$ and $O_{P_0}(1/\sqrt{n})$, respectively. Result (9) follows straightforwardly from the consistency of $H_n$ as an estimator of $H_0$, and the techniques (integration by parts) used in our proofs. Now, the third term on the right hand side can be written as:

$$\begin{aligned}
&-\int \frac{\int b(z,s) f_2(z,y,u,s;\beta) d(H_n - H_0)(y,z)}{\int f_2(z,y,u,s;\beta) dH_0(y,z)} E[dN_{X^0}(u,\theta) dN_{Y^0}(s,\eta_0)] + \\
&\int \frac{\int b(z,s) f_2(z,y,u,s;\beta) dH_n(y,z)}{\int f_2(z,y,u,s;\beta) dH_n(y,z) \int f_2(z,y,u,s;\beta) dH_0(y,z)} \times \\
&\quad \int f_2(z,y,u,s;\beta) d(H_n - H_0)(y,z) E[dN_{X^0}(u,\theta) dN_{Y^0}(s,\eta_0)]
\end{aligned} \tag{27}$$

To control the asymptotic behavior of the empirical process terms of the form $\int f d(H_n - H)$ in (6), we assume that $\mathcal{F}_2$ and $\mathcal{F}_b$ are uniformly bounded $H_0$-Donsker classes (RC6). By Example 2.10.8 of van der Vaart and Wellner (1996), we then know that $\mathcal{F}_b \times \mathcal{F}_2$ is also a uniformly bounded $H_0$-Donsker class. Let

$$\mathcal{F}_2^* \equiv \{(z,y) \to f_2^*(z,y,u,s;\beta) = b(z,s) f_2(z,y,u,s;\beta) : u,s,\beta\}$$

Under the above conditions conditions, we then have that $\sup_{f_2 \in \mathcal{F}_2} \mid \int f_2 d(H_n - H_0) \mid$ and $\sup_{f_2^* \in \mathcal{F}_2^*} \mid \int f_2^* d(H_n - H_0) \mid$ are $O_{P_0}(1/\sqrt{n})$. To control the behavior of the denominator of (6), we assume that, with probability 1,

$$I(\delta = \xi = 1) \int f_2(z,y,\tilde{L}(X,\tilde{M}(Y,Z;\eta),Z;\theta),\tilde{M}(Y,Z;\eta);\beta) dH_0(y,z) > \delta$$

for some $\delta > 0$, uniformly in $\beta \in \boldsymbol{\beta}$ (RC6). It then follows by empirical process theory (van der Vaart and Wellner, 1996) that (6) can be expressed as

$$-\int \frac{\int b(z,s) f_2(z,y,u,s;\beta) d(H_n - H_0)(y,z)}{\int f_2(z,y,u,s;\beta) dH_0(y,z)} E[dN_{X^0}(u,\theta) dN_{Y^0}(s,\eta_0)] +$$
$$\int \frac{\int b(z,s) f_2(z,y,u,s;\beta) dH_0(y,z)}{\left\{\int f_2(z,y,u,s;\beta) dH_0(y,z)\right\}^2} \int f_2(z,y,u,s;\beta) d(H_n - H_0)(y,z) E[dN_{X^0}(u,\theta) dN_{Y^0}(s,\eta_0)] +$$
$$o_{P_0}(1/\sqrt{n}).$$

This proves the lemma.

# References


Brem, H., Mahaley, M., Vick, N., Black, K., Schold, S., Burger, P., Friedman, A., Ciric, I., Eller, T., Cozzens, J., and Kenealy, J. (1991). Interstitial chemotherapy with drug polymer implants for the treatment of recurrent gliomas. *Journal of Neurosurgery*, 74:441–446.

Brem, H., Piantodsi, S., Burger, P., Walker, M., Selker, R., Vick, N., Black, K., Sisti, M., Brem, S., Mohr, G., Muller, P., Morawetz, R., and Schold, S. (1995). Placebo-controlled trial of safety and efficacy of intraoperative controlled delivery by biodegradable polymers if chemotherapy for recurrent gliomas. *Lancet*, 345.

Chang, S.-H. (2000). A two-sample comparison for multiple ordered event data. *Biometrics*, 56:183–189.

Fine, J., Jiang, H., and Chappell, R. (2001). On semi-competing risks data. *Biometrika*, 88:907–919.

Ghosh, D. and Lin, D. (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics*, 59(4):877–885.

Joffe, M. (2001). Administrative and artifical censoring in censored regression models. *Statistics in Medicine*, 20:2287–2304.

Kalbfleisch, J. and Prentice, R. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

Lin, D., Robins, J., and Wei, L. (1996). Comparing two failure time distributions in the presence of dependence censoring. *Biometrika*, 83:381–393.

Lin, D. and Ying, Z. (2003). Semiparametric regression analysis of longitudinal data with informative drop-outs. *Biostatistics*, 4:385–398.

Louis, T. (1981). Non-parametric analysis of an accelerated failure time model. *Biometrika*, 68:381–390.

Matsui, S. (2004). Analysis of times to repeated events in two-arm randomized trials with noncompliance and dependent censoring. *Biometrics*, 60(965-976).

Newey, W. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In R, E. and D, M., editors, *Handbook of Econometrics*, volume 4, chapter 36. Elsevier Science, Amsterdam.

Peng, L. and Fine, J. (2006). Rank estimation of accelerated lifetime models with dependent censoring. *Journal of the American Statistical Association*, 101(475):1085–1093.

Robins, J. (1995a). An analytic method for randomized trials with informative censoring: Part i. *Lifetime Data Analysis*, 1:241–254.

Robins, J. (1995b). An analytic method for randomized trials with informative censoring: Part ii. *Lifetime Data Analysis*, 1:417–434.

Robins, J. and Greenland, S. (2000). Comment on "causal inference without counterfactuals" by a.p. dawid. *Journal of the American Statistical Association*, 95:477–482.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701.

Tsiatis, A. (1990). Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics*, 18:354–372.

Wei, L. and Gail, M. (1983). Nonparametric estimation for a scale-change with censored observations. *Journal of the American Statistical Association*, 78:382–388.

## Captions

Table 1: Prognostic Baseline Factors, Stratified by Treatment Group

Table 2: Results of Simulation Studies

Table 3: Regression Estimates and Standard Errors (Subscripted) from Bivariate Linear Regression Model Using SLR and PF.

Figure 1: The function $r(t, w; \eta_{0a})$ represents the quantile-quantile mapping function between these distributions of death.The distributions without [ ] are equal to the distributions with brackets under randomization.

Figure 2: The function $q(t, u, w; \theta_{0a})$ represents the quantile-quantile mapping function between distributions of progression. The distributions without [ ] are equal to the adjacent distributions with brackets under randomization and rank preservation

Table 1:

| Prognostic Factor | Carmustime Polymer ($n = 110$) | Placebo Polymer ($n = 109$) |
|---|---|---|
| Resection $> 75$ | 74.5% | 73.4% |
| Age | 48.1 | 47.3 |
| White | 90.9% | 91.7% |
| Karnofsky $> 70$ | 55.5% | 51.4% |
| Local Radiation | 74.5% | 79.8% |
| Previous Nitrosoureas | 49.1% | 44.0% |
| Active Tumor Histology | 93.6% | 90.8% |
| Glioblastoma Subtype | 69.1% | 66.1% |
| Anaplastic Astrocytoma Subtype | 12.7% | 13.8% |
| Oliodendroglioma Subtype | 13.6% | 17.4% |
| Other Subtype | 4.5% | 2.8% |

Table 2:

(a) Bivariate Exponential

| | Truth | M.C. Avg. | Avg. Est. S.E. | M.C. S.D. | Cov. Prob. |
|---|---|---|---|---|---|
| $\theta_1$ | 0.4 | 0.4005 | 0.3218 | 0.3048 | 94.3% |
| $\theta_2$ | -0.1 | -0.1014 | 0.1603 | 0.1541 | 95.9% |
| $\theta_3$ | -0.2 | -0.1928 | 0.3208 | 0.3110 | 94.8% |
| $\eta_1$ | 0.2 | 0.1972 | 0.1540 | 0.1551 | 94.4% |
| $\eta_2$ | -0.3 | -0.2990 | 0.0765 | 0.0753 | 94.0% |
| $\eta_3$ | -0.4 | -0.4024 | 0.1525 | 0.1538 | 94.7% |

(b) Bivariate Normal

| | Truth | M.C. Avg. | Avg. Est. S.E. | M.C. S.D. | Cov. Prob. |
|---|---|---|---|---|---|
| $\theta_1$ | 0.4 | 0.3874 | 0.2371 | 0.2344 | 94.7% |
| $\theta_2$ | -0.1 | -0.0951 | 0.1173 | 0.1150 | 94.7% |
| $\theta_3$ | -0.2 | -0.2095 | 0.2345 | 0.2413 | 94.9% |
| $\eta_1$ | 0.2 | 0.1992 | 0.1518 | 0.1516 | 94.4% |
| $\eta_2$ | -0.3 | -0.2966 | 0.0741 | 0.0738 | 93.8% |
| $\eta_3$ | -0.4 | -0.3963 | 0.1500 | 0.1475 | 94.8% |

Table 3:

| | Cerebellar Progression | | Death |
|---|---|---|---|
| Covariate | SRL | PF | SLR/PF |
| Treatment | $0.377_{0.155}$ | $0.345_{0.293}$ | $0.280_{0.111}$ |
| Resection $> 75$ | $0.446_{0.181}$ | $0.354_{0.272}$ | $0.339_{0.130}$ |
| Age (in decades) | $-0.270_{0.074}$ | $-0.214_{0.200}$ | $-0.172_{0.047}$ |
| White | $-0.212_{0.281}$ | $-0.397_{0.319}$ | $-0.421_{0.206}$ |
| Karnofsky $> 70$ | $0.075_{0.170}$ | $0.064_{0.330}$ | $0.292_{0.120}$ |
| Local Radiation | $0.466_{0.174}$ | $0.244_{0.402}$ | $0.399_{0.120}$ |
| Previous Nitrosoureas | $-0.606_{0.176}$ | $-0.444_{0.367}$ | $-0.274_{0.121}$ |
| Active Tumor Histology | $-0.152_{0.343}$ | $0.030_{0.291}$ | $-0.458_{0.283}$ |
| Anaplastic Astrocytoma Subtype | $1.022_{0.285}$ | $0.606_{0.409}$ | $0.471_{0.152}$ |
| Oliodendroglioma Subtype | $1.204_{0.323}$ | $1.132_{0.405}$ | $0.839_{0.230}$ |
| Other Subtype | $0.899_{0.590}$ | $0.852_{0.316}$ | $0.816_{0.276}$ |

Figure 1:

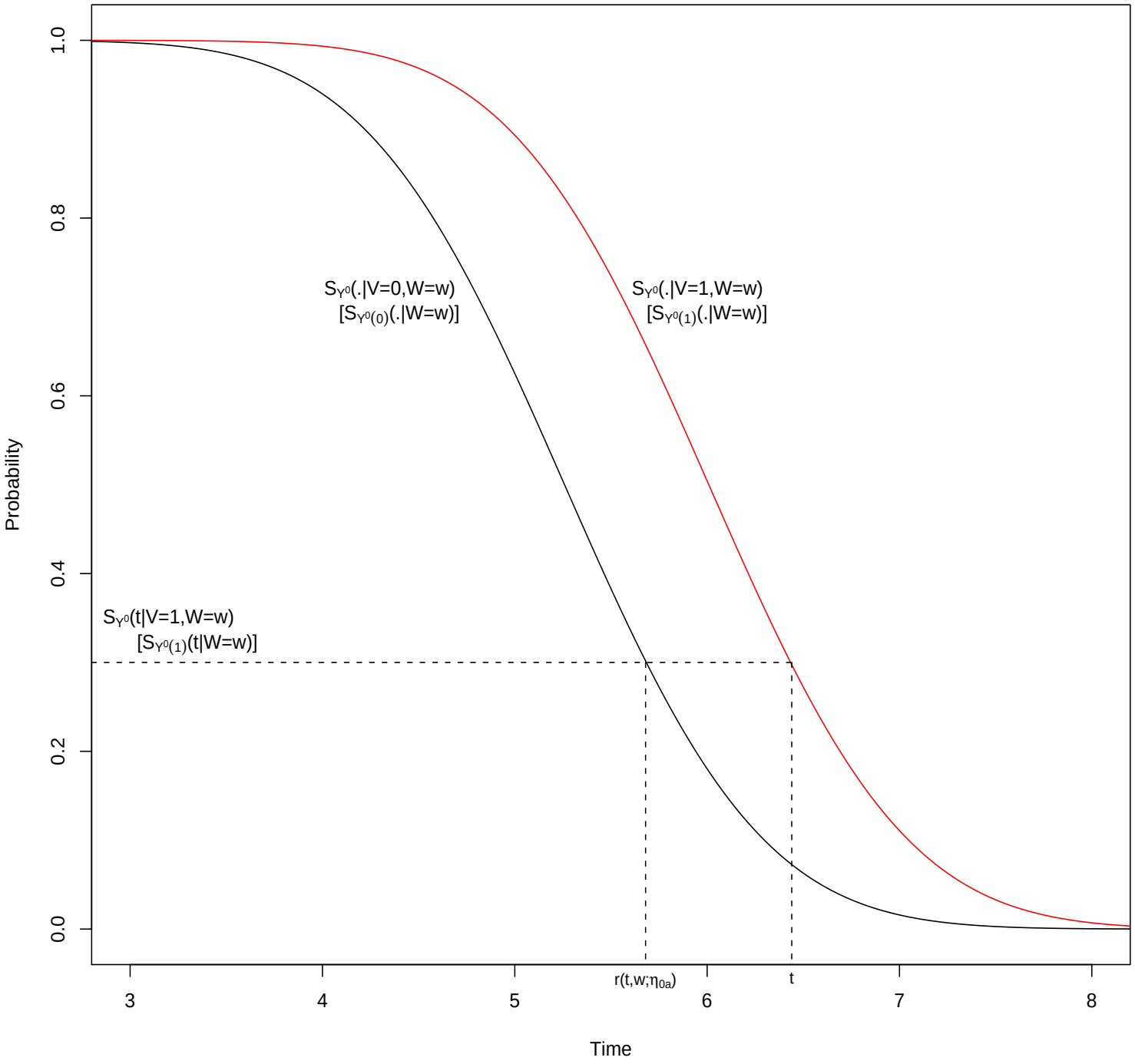

Figure 2: