



---

UW Biostatistics Working Paper Series

---

3-16-2010

# Panel Count Data Regression with Informative Observation Times

Petra Buzkova

*University of Washington*, [buzkova@u.washington.edu](mailto:buzkova@u.washington.edu)

---

## Suggested Citation

Buzkova, Petra, "Panel Count Data Regression with Informative Observation Times" (March 2010). *UW Biostatistics Working Paper Series*. Working Paper 362.  
<http://biostats.bepress.com/uwbiostat/paper362>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

# Panel Count Data Regression with Informative Observation Times

Petra Bůžková

Department of Biostatistics, University of Washington, Seattle, WA 98195

email: buzkova@u.washington.edu

## Abstract

When patients are monitored for potentially recurrent events such as infections or tumor metastases, it is common for clinicians to ask patients to come back sooner for follow-up based on the results of the most recent exam. This means that subjects' observation times will be irregular and related to subject-specific factors. Previously proposed methods for handling such panel count data assume that the dependence between the events process and the observation time process is time-invariant. This article considers situations where the observation times are predicted by time-varying factors, such as the outcome observed at the last visit or cumulative exposure. Using a joint modeling approach, we propose a class of inverse-intensity-rate-ratio weighted estimators that are root  $n$  consistent and asymptotically normal. The proposed estimators use estimating equations and are fairly simple and easy to compute. We demonstrate the performance of the method using simulated data and illustrate the approach using a cancer study dataset.

*Keywords:* Informative follow-up; Panel count data; Recurrent events; Semi-parametric regression; Time-varying covariates.



# 1 Introduction

Recurrent events data arise when a certain event can occur repeatedly over time. Some studies are able to monitor study patients continuously, recording the times of all event occurrences. Such data are called event history data. However, it may not be feasible to monitor patients continuously for events. Therefore, a common study design is to observe patients only at discrete time points. Only the counts of events up to the observation time are known, with the exact event times unknown. The available data are interval-censored recurrent events data and are often referred to as “panel count data”. As an example, consider a study of bladder cancer conducted by the Veterans Administration Cooperative Urological Research Group (Byar 1980). The bladder tumor occurrences were observed not continuously over time, but only at each follow-up visit when they were counted, measured and removed.

With panel count data two processes arise: one process for the recurrent events, that is not fully observed but of primary interest, and a second process for observation times. If observation time points are not fixed across individuals and they vary from subject to subject in a non-random way, it is important to characterize correctly the mutual relationship of those processes to draw correct inference about the association parameters in the model for the mean cumulative number of events.

Several panel count data methodologies have been suggested under the assumption that the recurrent events process and the observation times process are independent. This scenario is analogous to missingness completely

at random (Little & Rubin 2002), including situations in which the observation times are fixed in advance. For example, Sun & Kalbfleisch (1995) and Wellner & Zhang (2000) proposed non-parametric estimators of the mean function of a counting process with panel count data independent of covariates. Zhang (2002) and Wellner & Zhang (2007) extended the above methods for regression settings allowing the mean function of the cumulative number of events to dependent on covariates. They focused on estimating both the contrast parameters and the baseline mean function, assuming homogeneous and non-homogeneous Poisson processes.

A few analyses were suggested for the situation when the recurrent events process and the observation times process are dependent on covariates of the recurrent events model. This is analogous to missingness at random. Focusing on estimation of the contrast parameters, Sun & Wei (2000) suggested a semiparametric approach, which was further generalized by Hu, et al. (2003). They analyzed the same bladder cancer study and suggested that visit patterns were different from subject to subject. They allowed for dependence on treatment assignment and initial condition such as number of tumors observed at the beginning of the study and the size of the largest initial tumor, all covariates of the model for the mean cumulative number of tumors.

Huang, et al. (2006) were the first to propose a methodology for panel count data with informative observation times, that is a situation where the recurrent events process and the observation times process are dependent conditional on covariates of the model for the cumulative number of

events. They studied nonparametric and semiparametric models where the conditional dependence is accounted for by the existence of a frailty variable that is multiplicative in the mean model for cumulative number of events. They use conditional likelihood maximization and estimating equations to draw inference. Sun, et al. (2007) used a joint modeling approach for the mean cumulative number of recurrent events and the observation times process, also with a multiplicative latent variable that should accommodate the correlation between observation times and recurrent events. Using an EM algorithm, He, et al. (2009) proposed a similar approach of joint modeling with latent variables.

All three methodologies address the dependence between the events process and the observation time process by including a time-fixed latent variable into the models. This addresses dependence induced by unmeasured baseline variables. However, none of these methods accommodate situations where the dependence between the two processes is induced by time-varying factors. For example, a clinician may request that a patient returns for a follow-up visit sooner or later based on the most recent exam. In the bladder cancer study we argue that the observation times depend on the number of new tumors at the last visit. If the number of new tumors was large, a clinician asks the patient to come back earlier than he would otherwise. Generally, the observation times can be predicted by past observed number of recurrent events, cumulative exposure, or other factors related to the past observed outcome.

In Section 2 we introduce our notation and our proposed models for the mean cumulative number of events and for the observation times. Both models are semiparametric models that are often used for modeling recurrent events data. We propose an inference approach in Section 3. To adjust for the dependence between the events process and the observation time process we utilize an inverse-intensity rate ratio (IIRR) weighted approach, introduced in Bůžková & Lumley (2007) for generalized linear models and extended to semiparametric linear and log-linear models in Bůžková & Lumley (2009) and Bůžková & Lumley (2008), respectively. The proposed methodology, offering a flexible modeling approach, also offers an estimation approach that is fairly simple in formulas and enumeration. In Section 4 we report results from a simulation study of panel count data observed at independent, dependent and informative observation times to demonstrate the performance of the estimator under finite sample size. We illustrate our method by applying it to the bladder tumor study in Section 5. The treatment effect was increased when accounting for dependence of visit times on the number of tumors observed at the last follow-up visit. We conclude this article with a discussion section. The simulations and data analysis were carried out in R (R Development Core Team 2007).



## 2 Preliminaries and notation

### 2.1 Models and assumptions

Consider a study with  $n$  subjects who may experience recurrent events of interest. For each subject  $i$  let  $N_i(t)$  denote the cumulative number of events before time  $t$ . Suppose that  $N_i = \{N_i(t) : t \geq 0\}$  is a counting process with conditional mean function given by

$$E[N_i(t)|X_i] = \exp\{\beta_0^T X_i\} \Lambda_0(t), \quad (1)$$

where  $X_i$  is a time-independent covariate with  $p$  components and  $\Lambda_0(\cdot)$  is an unspecified baseline mean function of the cumulative number of events and is considered a nuisance parameter. The goal is to make inference about the association parameter  $\beta_0$ .

For subject  $i$ , suppose that the counting process  $N_i$  is observed only at  $K_i$  discrete time points  $T_{i1}, T_{i2}, \dots, T_{iK_i}$  with  $0 \equiv T_{i0} < T_{i1} < T_{i2} < \dots < T_{iK_i} \leq \tau$  where  $K_i$  is an integer-valued random variable and  $\tau \in (0, \infty)$  is the end of study time. Let  $C_i, C_i \leq \tau$ , denote the censoring time after which further follow-up of individual  $i$  is impossible, and let  $\tilde{N}_i(t) = \sum_{k=1}^{K_i} I(T_{ik} \leq t)$  be the observed counting process of the number of observations of individual  $i$  by time  $t$ . The underlying uncensored counting process is  $\tilde{N}_i^* = \{\tilde{N}_i^*(t) : t \geq 0\}$  with  $\tilde{N}_i(t) = \tilde{N}_i^*(t \wedge C_i)$ , where  $a \wedge b$  stands for  $\min(a, b)$ . We consider a conditional rate model for the uncensored observation times

$$E[d\tilde{N}_i^*(t)|Z_i(t)] = \exp\{\gamma_0^T Z_i(t)\} d\tilde{\Lambda}_0(t). \quad (2)$$

$Z_i(\cdot)$  has  $g$  components that can be time-dependent and  $\tilde{\Lambda}_0(\cdot)$  is the cumulative baseline intensity. Model (2) can accommodate a mixture of predefined observation times and random observation times that depend on covariates  $Z_i(\cdot)$ . We write  $\xi_i(t) = I(C_i > t)$  for the at-risk process.

Both models (1) and (2) are very common for regression analysis of recurrent events data and have been used previously by many authors. For example, Zhang (2002) and Wellner & Zhang (2007) considered model (1) for panel count data assuming independent observation times, Sun & Wei (2000) and Hu et al. (2003) in their methods used model (1) as well as model (2) with time-invariant predictors  $Z_i = X_i$ . The methodology of Huang et al. (2006) multiplicatively added a latent variable into model (1) and Sun et al. (2007) and He et al. (2009) multiplicatively added latent variables into both of the models.

We require two assumptions. First, we assume non-informative censoring for the mean of the outcome,

$$E[N_i(t)|X_i, C_i \geq t] = E[N_i(t)|X_i]. \quad (3)$$

This assumption says that the mean cumulative number of events  $N_i(t)$  depends on covariates  $X_i$  and censoring  $C_i$  through covariates  $X_i$  only. Second, we assume that observation times depend on covariates  $X_i$ , on outcome  $N_i(t)$  and censoring  $C_i$  only through covariates  $Z_i(t)$ , i.e.

$$E[d\tilde{N}_i^*(t)|Z_i(t), X_i, N_i(t), C_i \geq t] = E[d\tilde{N}_i^*(t)|Z_i(t)]. \quad (4)$$

Because  $Z_i(t)$  has the appealing ability to include the past observed outcome,



variables in the causal pathway between the covariates  $X_i$  and the outcome, or auxiliary factors related with outcome we consider this assumption weak and natural.  $Z_i(t)$  can not however encompass the current outcome  $N_i(t)$ .

For the analysis of recurrent events, it is more convenient to work directly with the mean function defined in (1) rather than the rate function due to the incomplete nature of the observed information. On the other hand, the model for observation times can be the more flexible proportional rate model (2) naturally incorporating time-varying predictors as that process is fully observed up to a censoring time.

## 2.2 Weights and estimation in the observation time model

To adjust for the informative observation times we use inverse intensity rate ratio weighting (Bůžková & Lumley 2007). For the  $i$ -th individual at time  $t$  define inverse weights

$$\rho_i(t; \gamma_0, \delta) = \frac{\exp\{\gamma_0^T Z_i(t)\}}{\exp\{\delta^T X_i\}}, \quad (5)$$

where  $\gamma_0$  is the true parameter vector in the model for observation times (2) and  $\delta$  is any fixed vector of length  $p$ . The true  $\gamma_0$  and thus the true weights are unknown but can be estimated from the data. The best choice of  $\delta$  we base on an estimator of  $\delta_0$ , the contrast parameter vector in a proportional rate model analogous to (2) but conditioning on covariates  $X_i$  alone. We use this from now on. The numerator of the stabilized inverse weights accounts for the prediction of observation times by  $Z_i(t)$  and the denominator accounts

for the prediction of observation times by  $X_i$  alone. Another natural choice for  $\delta$  is to set it to zero, simplifying the definition of the inverse weight but losing some convenience further in the estimation method.

The estimation of the parameter vector  $\gamma_0$  in the model for observation times (2) is described in detail in Lin, et al. (2000). We briefly introduce it here because we use it to estimate the inverse weights. The estimation in the model for observation times is based on a random process

$$\tilde{\mathcal{M}}_i(t; \gamma, \tilde{\Lambda}) = \tilde{N}_i(t) - \int_0^t \xi_i(s) \exp\{\gamma^T Z_i(s)\} d\tilde{\Lambda}(s), \quad (6)$$

that has mean zero at  $\gamma_0, \tilde{\Lambda}_0(\cdot)$ . Define the mean  $Av_1$  at time  $t$  of any variable  $Z(\cdot)$  given parameter  $\gamma$  as

$$Av_1(Z)(t; \gamma) = \sum_{i=1}^n Z_i(t) \frac{\xi_i(t) \exp\{\gamma^T Z_i(t)\}}{\sum_{j=1}^n \xi_j(t) \exp\{\gamma^T Z_j(t)\}}. \quad (7)$$

The parameter  $\gamma_0$  is consistently estimated by a solution to the estimating equation  $\tilde{U}(\hat{\gamma}) = 0$ , where

$$\tilde{U}(\gamma) = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - Av_1(Z)(t; \gamma)\} d\tilde{N}_i(t). \quad (8)$$

The estimating function (8) is the same as for Cox proportional hazards models, however not the variance of  $\hat{\gamma}$ . Denote  $V^{\otimes 2} = VV^T$  and define the Aalen–Breslow estimator of  $\tilde{\Lambda}_0(t)$  as  $\hat{\Lambda}(t) = \sum_{i=1}^n \int_0^t \left[ d\tilde{N}_i(s) / \sum_{j=1}^n \xi_j(s) \exp\{\hat{\gamma}^T Z_j(s)\} \right]$ . The asymptotic variance of  $\sqrt{n}(\hat{\gamma} - \gamma_0)$  is consistently estimated by  $\hat{A}^{-1} \hat{\Sigma} \hat{A}^{-1}$ , where

$$\begin{aligned} \hat{A} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau [Z_i(t) - Av_1(Z)(t; \hat{\gamma})]^{\otimes 2} \xi_i(t) \exp\{\hat{\gamma}^T Z_i(t)\} d\hat{\Lambda}(t) \quad (9) \\ \hat{\Sigma} &= \frac{1}{n} \sum_{i=1}^n \left[ \int_0^\tau [Z_i(t) - Av_1(Z)(t; \hat{\gamma})] d\tilde{\mathcal{M}}_i(t; \hat{\gamma}, \hat{\Lambda}) \right]^{\otimes 2}. \end{aligned}$$

### 3 Estimation in the recurrent events model

In this section we develop a class of estimators for panel count data when the recurrent events process and the observation times process are dependent conditional on covariates of the mean model for cumulative number of events. We use the observation time model (2), allowing the observation times to depend on predictors  $Z_i(t)$ . Let us define a process  $\{\mathcal{M}_i(t), t \in [0, \tau]\}$

$$\begin{aligned} \mathcal{M}_i(t; \beta, \gamma, \delta, \mathcal{A}) &= \int_0^t \frac{1}{\rho_i(s; \gamma, \delta)} \left[ N_i(s) d\tilde{N}_i(s) \right. \\ &\quad \left. - \exp\{\beta^T X_i\} \xi_i(s) \exp\{\gamma^T Z_i(s)\} d\mathcal{A}(s) \right] \\ &= \int_0^t \left[ \frac{N_i(s) d\tilde{N}_i(s)}{\rho_i(s; \gamma, \delta)} - \xi_i(s) \exp\{(\beta + \delta)^T X_i\} d\mathcal{A}(s) \right], \end{aligned}$$

where the inverse weights  $\rho_i(\cdot)$  were introduced in (5) and  $\mathcal{A}(t) = \int_0^t \Lambda(s) d\tilde{\Lambda}(s)$ . Define  $\mathcal{A}_0(t) = \int_0^t \Lambda_0(s) d\tilde{\Lambda}_0(s)$ . For any fixed  $\delta$   $E[d\mathcal{M}_i(t; \beta_0, \gamma_0, \delta, \mathcal{A}_0) | X_i] = 0$ , as shown in the Appendix. A natural set of equations for estimators of  $\mathcal{A}_0(\cdot)$  and  $\beta_0$  is

$$\sum_{i=1}^n \mathcal{M}_i(t; \beta, \gamma_0, \delta, \mathcal{A}) = 0 \quad \forall t \in [0, \tau] \quad (10)$$

$$\sum_{i=1}^n \int_0^\tau w(t) X_i d\mathcal{M}_i(t; \beta, \gamma_0, \delta, \mathcal{A}) = 0. \quad (11)$$

Equation (10) at time  $t \in [0, \tau]$  and equation (11) yield

$$\begin{aligned} \hat{\mathcal{A}}(t) &= \sum_{i=1}^n \int_0^t \left[ \frac{N_i(s) d\tilde{N}_i(s)}{\rho_i(s; \gamma_0, \delta)} / \sum_{j=1}^n \xi_j(s) \exp\{(\hat{\beta} + \delta)^T X_j\} \right] \\ U(\beta; \gamma_0, \delta) &= \sum_{i=1}^n \int_0^\tau \frac{w(t)}{\rho_i(t; \gamma_0, \delta)} \{X_i - Av_2(X)(t; \beta, \delta)\} N_i(t) d\tilde{N}_i(t) \quad (12) \end{aligned}$$

where the mean  $Av_2(\cdot)$  is defined as

$$Av_2(X)(t; \beta, \delta) = \sum_{i=1}^n X_i \frac{\xi_i(t) \exp\{(\beta + \delta)^T X_i\}}{\sum_{j=1}^n \xi_j(t) \exp\{(\beta + \delta)^T X_j\}}. \quad (13)$$

We show in the Appendix that the parameter vector  $\beta_0$  of length  $p$  can be consistently estimated by  $\hat{\beta}$ , the solution to a set of estimating equations  $U(\hat{\beta}; \hat{\gamma}, \delta) = 0$ . Further,  $\sqrt{n}(\hat{\beta}(\hat{\gamma}, \delta) - \beta_0)$  converges in distribution to a zero-mean Normal random vector with a covariance matrix consistently estimated by  $\hat{D}^{-1} \hat{V} \hat{D}^{-1}$ , where

$$\begin{aligned} \hat{D} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{w(t)}{\rho_i(t; \hat{\gamma}, \delta)} \left[ X_i - Av_2(X)(t; \hat{\beta}, \delta) \right]^{\otimes 2} \xi_i(t) \exp\{(\hat{\beta} + \delta)^T X_i\} d\hat{A}(t) \\ \hat{V} &= \frac{1}{n} \sum_{i=1}^n \left[ \int_0^\tau \frac{w(t)}{\rho_i(t; \hat{\gamma}, \delta)} \left[ X_i - Av_2(X)(t; \hat{\beta}, \delta) \right] d\mathcal{M}_i(t; \hat{\beta}, \hat{\gamma}, \delta, \hat{A}) \right. \\ &\quad \left. - \hat{H} \hat{A}^{-1} \int_0^\tau [Z_i(t) - Av_1(Z)(t; \hat{\gamma})] d\tilde{\mathcal{M}}_i(t, \hat{\gamma}, \hat{\Lambda}(\cdot)) \right]^{\otimes 2}, \end{aligned}$$

and  $\hat{H} = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{w(t)}{\rho_i(t; \hat{\gamma}, \delta)} \left[ X_i - Av_2(X)(t; \hat{\beta}, \delta) \right] Z_i(t)^T N_i(t) d\tilde{N}_i(t)$ . The process  $\{\tilde{\mathcal{M}}(t), t \in [0, \tau]\}$  defined in equation (6) and the matrix  $\hat{A}$  defined in (9) are used in the solution of the observation time model (2). The second term in the squared brackets of the covariance matrix  $V$  accounts for the fact that  $\gamma_0$  is estimated rather than known. Bootstrapping is a good alternative for estimating the variance in situations with stable weights. Because the estimation approach above is valid for any fixed  $\delta$ , we can substitute it by the estimator  $\hat{\delta}$  based on an observation time model with covariates  $X$ . We note that we do not require a consistent estimator of the nuisance baseline cumulative intensity  $\Lambda_0(t)$  of the recurrent events process nor the baseline

intensity  $d\tilde{\Lambda}_0(t)$  of the observation times process and thus do not need to use any smoothing techniques for achieving consistency.

In a situation where observation times are independent of any covariates, that is  $\gamma_0 = 0$  and  $\delta_0 = 0$ , the estimating function (12) simplifies to

$$U^\dagger(\beta) = \sum_{i=1}^n \int_0^\tau w(t) \{X_i - Av_1(X)(t; \beta)\} N_i(t) d\tilde{N}_i(t). \quad (14)$$

We call a solution to (14) an independent estimator. Similarly, an estimator for a situation where observation times are dependent only on the covariates of the recurrent events model with  $Z_i(t) = X_i$  can be computed using the estimation function

$$U^\ddagger(\beta; \delta_0) = \sum_{i=1}^n \int_0^\tau w(t) \{X_i - Av_2(X)(t; \beta, \delta_0)\} N_i(t) d\tilde{N}_i(t). \quad (15)$$

Again this is a simplification of the the estimating function (12) that provides the IRR weighted estimator. It is equivalent to the estimation function (13) of Hu et al. (2003).

## 4 Simulation study

We conducted a simulation study to examine the finite sample behavior of the proposed estimators under different scenarios. We consider different relationships between the recurrent events process and the observation times process.

We fitted an outcome model with a single time-independent covariate  $X_i$

$$E[N_i(t)|X_i] = \exp\{\beta_{01}X_i\}\Lambda_0(t).$$

We are interested in drawing inference about the parameter  $\beta_{01}$ , which summarizes the association between the covariate  $X$  and the mean number of recurrent events. In contrast, the observation times are fitted using two covariates  $Z_i(t) = (X_i, Z_{2i}(t))$ . We take  $Z_{2i}(t)$  to be dependent on  $X_i$  and also generate the recurrent events data using both  $X_i$  and  $Z_{2i}(t)$ . At the same time we make sure that the marginal model for mean outcome is correct. Thus, the observation times process and the outcome process are dependent, conditioning on the covariate  $X_i$  of the outcome model. The dependence is induced by  $Z_{2i}(t)$ , which is associated with the  $X_i$  as well as the outcome  $N_i(t)$ .

For each individual, we generated data by the following scheme: Let  $X_i$  be Bernoulli(0.5), representing a treatment or control assignment. Covariate  $Z_{2i}(t)|X_i \sim X_i N(0, 1) + (1 - X_i)N(1, 1)$  are *i.i.d.* random variables. The effect of the treatment is to lower  $Z_{2i}(t)$ .  $Z_{2i}(\cdot)$  is a step function that can change its value at the observation times  $\{T_{i1}, T_{i2}, \dots, T_{iK_i}\}$ , so that  $Z_{2i}(t) = Z_{2i}(T_{ij}); T_{ij} \leq t < T_{ij+1}$ . The outcome was generated using a random effect semiparametric Poisson-Gamma model

$$E [N_i(t)|X_i, Z_{2i}(t), \phi_i] = \frac{\exp\{\beta_{01}X_i + \beta_{02}Z_{2i}(t)\}}{E [\exp\{\beta_{02}Z_{2i}(t)\}|X_i]} \phi_i \Lambda_0(t)$$

with the true parameter values  $\beta_{01} = 1$ ,  $\beta_{02} = -0.4$ . The independent random effect  $\phi_i$  are  $\Gamma$  distributed with mean  $\mu_\phi = 1$  and variance  $\sigma_\phi^2 = 0.01$ . We note that  $E [\exp\{\beta V\}]$  has a closed form for a random variable  $V \sim N(\mu, \sigma^2)$ , equal to  $\exp\{\beta\mu + \beta^2\sigma^2/2\}$ . The observation times follow a random effect Poisson counting process with intensity  $\lambda_i(t) = \eta_i \exp\{\gamma_{01}X_{1i}(t) + \gamma_{02}Z_{2i}(t)\}$ ,

where random variables  $\eta_i$  are  $\Gamma$  distributed with mean  $\mu_\eta = 1$  and variance  $\sigma_\eta^2 = 0.01$  independent of any other variable. For each individual the observation times are positively correlated, unless  $\sigma_\eta = 0$ . Observation times are generated by utilizing the property that time between observations in a Poisson counting process has an Exponential distribution with parameter  $\lambda_i(t)$ . We study the estimators under three options of the parameter  $\gamma_0 = (\gamma_{01}, \gamma_{02})$ . The informative observation times scenario is when  $\gamma_{01} = -0.6, \gamma_{02} = 0.8$ , the dependent observation times scenario is  $\gamma_{01} = -0.6, \gamma_{02} = 0$  and finally the independent scenario is  $\gamma_{01} = \gamma_{02} = 0$ . Censoring variables  $C_i$  are uniform on  $(\tau/2, \tau)$ , with  $\tau$  set to 3. Both required assumptions (3) and (4) are met. We considered sample sizes  $n = 50, 100, 200$  and 500. We ran 1000 simulations at each combination of parameters.

Tables 1 and 2 provide summaries of the estimates of  $\beta_{01}$  assuming that the observation times are covariate independent (independent estimator (14)), that they are associated with covariate  $X$  only (dependent estimator (15)) and both  $X$  and  $Z(t)$  (IIRR weighted estimator (12)). Using the independent estimator we see that it correctly estimates  $\beta_{01}$  in the situation when the observation times are truly independent. Using the dependent estimator we see that it correctly estimates  $\beta_{01}$  in the situation when the observation times are either independent or dependent on  $X$ . Biases of the estimates are negligible and sampling standard errors (SSEs) are close to the mean of bootstrapped standard errors (BSEs) even for the sample size of 50. The 95% coverage probability is ranging from 94-97%. As expected, the inde-

pendent estimator fails when observation times are dependent on covariate  $X$  and also when they are informative, that is dependent on  $X$  and  $Z_2(t)$ . The dependent estimator fails when observation times are informative, that is dependent on  $X$  as well as  $Z_2(t)$ . Biases of the estimates and mean squared errors (MSEs) are large and the coverage probability is very small. Finally, the IIRR weighted estimator works well in all three scenarios of observation times. The estimates do not seem to lose efficiency compared to the independent estimates when observation times are independent on  $X$  and  $Z_2(t)$ . They also do not lose efficiency when observations are dependent on  $X$  alone. We note that the bootstrapped standard error is usually slightly bigger than the sampling standard error in all studied situations, slightly overestimating the true standard error of the estimate of  $\beta_{01}$ .

## 5 Example

We analyze the bladder cancer data extracted from Andrews & Herzberg (1985, p. 253-260) and conducted by the Veterans Administration Cooperative Urological Research Group. The study, a randomized clinical trial of three treatments, placebo (n=47), pyridoxine pills (n=31) and thiotepa instillation into the bladder (n=38), was conducted for patients with superficial bladder tumor when entering the trial. At each follow-up visit tumors were counted, measured and then removed if observed, and the treatment was continued. The treatment effects on suppressing the recurrence of bladder tumor, especially the thiotepa instillation, have been explored by many au-



thors, for example, Wellner & Zhang (2007), Sun & Wei (2000) and Hu et al. (2003), assuming that observation times are not informative about the mean cumulative number of tumors. The entire dataset for placebo and thiotepa groups is shown in Tables 1 and 2 of Hu et al. (2003), demonstrating that the visiting patterns were different from subject to subject. During the 48 month study time the number of visits among the 85 placebo and thiotepa patients ranged from 1 to 38 with median of 8. Sun & Wei (2000) and Hu et al. (2003) suggested that observation times may be dependent on variables and their method enabled them to include the set of baseline covariates from the mean model for cumulative number of tumors.

We investigated the effect of thiotepa on mean cumulative number of tumors under various degrees of complexity of the relationship of the cumulative number of tumors and the observation times process; first assuming the observation times to be independent of the covariates; second assuming dependence on the covariates of the recurrent events model; and third allowing for dependence on an additional covariate that is time-varying. As such we used the number of new tumors at the last visit - a variable that was intentionally not included among covariates in the model for recurrent events. We emphasize that in a clinical setting it is often the case that a clinician will ask a patient to come back earlier based on the patients performance so far. We suggest that the observation times depend on the number of new tumors at the last visit. If the number of new tumors was large, a clinician is likely to ask the patient to come back earlier than he would otherwise.

The proposed multiplicative mean model for the number of recurrent events that was used in previous analyses of this dataset is

$$E[N(t)] = \exp\{\beta_{01}X_1 + \beta_{02}X_2 + \beta_{03}X_3\} \Lambda_0(t), \quad (16)$$

where  $X_1$  represents the dummy variable for the treatment of thiotepa instillation, and  $X_2$  and  $X_3$  represent the count of bladder tumors at the beginning of the study and the size of the largest tumor at the beginning of the study, respectively.

Table 3 shows the results from the three suggested methods. Using the independent estimator, assuming no dependence of observation times on covariates, we obtained  $\hat{\beta} = (-1.30, 0.24, -0.05)$ . We calculated the sampling standard errors of the estimates from 500 bootstrap resamples of the dataset. Next, we computed the association estimates assuming that the observation times depend on the covariates of the outcome model (16). We obtained  $\hat{\gamma} = (0.47, -0.01, 0.05)$  with  $(0.17, 0.05, 0.06)$  the estimates of the standard errors of  $\hat{\gamma}$ . We obtain  $\hat{\beta} = (-1.49, 0.24, -0.10)$ . We calculated the sampling standard errors of the estimates from 500 bootstrap resamples of the dataset, for each resample recalculating the estimates of  $\gamma_0$ . Our findings when assuming both independence and dependence on covariates of the recurrent events model are consistent with the analysis results presented by Hu et al. (2003).

Last, we computed the association estimates when allowing for informative observation times. We propose the observation time model  $E[d\tilde{N}^*(t)] = \exp\{\gamma_{01}Z_1 + \gamma_{02}Z_2 + \gamma_{03}Z_3 + \gamma_{04}Z_4(t)\} d\tilde{\Lambda}_0(t)$ , where the predictors of obser-

vation times consist of the three covariates from the model for recurrent events  $X_1, X_2, X_3$  and the number of tumors at the last visit as the last predictor  $Z_4(t)$  which is time-varying. We obtained  $\hat{\gamma} = (0.46, -0.01, 0.05, 0.17)$  with estimated standard errors  $(0.16, 0.04, 0.06, 0.02)$ . Indeed, the hazard of a next visit is elevated with a higher count of tumors observed at last visit. The IIRR weights based on  $\hat{\delta}$  and  $\hat{\gamma}$  ranged from 0.88 up to 1.61. Setting the estimating function (12) to zero, we obtain  $\hat{\beta} = (-1.75, 0.24, -0.10)$ . Again, we calculated the standard errors of the estimates from 500 bootstrap resamples of the dataset, for each resample recalculating the estimates of  $\gamma_0$  and  $\delta_0$ .

All three estimates of  $\beta_{01}$ , the coefficient of the indicator for thiotepa treatment, suggest that thiotepa reduces the mean number of tumors. All three Wald tests indicate statistical significance. Comparing the three methods, the negative coefficient increases in magnitude. This suggests that the effect of thiotepa may be underestimated when ignoring the dependence of observation times on covariates, both covariates of the recurrent event model and the additional covariate of number of tumors at the last visit. The ratio of expected tumor counts increases from 3.66 comparing placebo to thiotepa groups by the independent method to 4.44 by the dependent method and to 5.75 by the IIRR weighted method. The ratio is 1.6 times higher comparing the IIRR weighted method and the independent method.

## 6 Discussion

In this article we considered regression analysis of recurrent events when observation times are informative, focusing on estimation of the association parameter  $\beta_0$  in the semiparametric marginal model  $E[N_i(t)|X_i] = \exp\{\beta_0^T X_i\}\Lambda_0(t)$  when panel count data are available. The proposed methodology allows for the predictors of the observation times to be time-varying. We proposed a two step procedure, estimating first IIRR weights from a proportional rate model for observation times and second estimating the association parameters of interest based on a proportional mean model above. The estimating approach is easy to implement.

A key advantage of the proposed methodology over currently available approaches using latent variables is that the informativeness of observation times can vary for each individual over time; the predictors in the observation time model can be time-varying. They can include past observed outcome, or variables in the causal pathway between exposure and outcome. By modeling observation times with real variables, unlike with unknown latent variables, they need to be measured. As a consequence, researchers must consider the observation time process carefully when planning a study in order to identify and correctly measure the predictors of the observation times. This can be viewed as both an advantage and a drawback. It is an advantage because knowledge is gained about the parameters in the observation time model, describing the observation times mechanism better. It is a drawback because the predictors of the observation time need to be measured. The values of

the time-varying predictors at all observation times are needed. When this is not realistic, using the most recently measured value can be a sensible approximation.

For drawing inference about  $\beta$  we recommend using the bootstrap procedure as in the example. Our simulations consistently demonstrated very good properties for the bootstrap. We have computed the model-based formula for the variance estimator of the IIRR weighted estimator as well. It behaved well for sample sizes as small as  $n = 50$ . Not to detract a reader from the main line of the paper, we did not present the formulas for the variance estimators for the independent and dependent estimators, which are simplifications of the formula for the variance estimator of the IIRR weighted estimator. Therefore, we did not show the results we obtained in our simulation study.

We illustrated our estimation approach on an example of tumor data. Other scenarios where the IIRR weighted panel count data approach could be used are epileptic seizures, flare ups, migraines, infections or rescue medications.

## References

- D. F. Andrews & A. M. Herzberg (1985). *Data: A collection of problems from many fields for the student and research worker*. Springer, New York.
- P. Bůžková & T. Lumley (2007). ‘Longitudinal data analysis for generalized linear models with follow-up dependent on outcome-related variables’. *Canadian Journal of Statistics* **35**:485–500.

- P. Bůžková & T. Lumley (2008). ‘Semiparametric Loglinear regression for longitudinal measurements subject to irregular, biased follow-up’. *Journal of Statistical Planning and Inference* **138**:2450–2461.
- P. Bůžková & T. Lumley (2009). ‘Semiparametric modeling of repeated measurements under outcome-dependent follow-up’. *Statistics in Medicine* **28**:987–1003.
- D. P. Byar (1980). ‘The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparison of placebo, pyridoxine, and topical thiotepa’. In *Bladder tumors and other topics in urological oncology*, pp. 363–370. Plenum, New York, eds M. Pavone-Macaluso, P. H. Smith and F. Edsmyn.
- X. He, et al. (2009). ‘Semiparametric analysis of panel count data with correlated observation and follow-up times’. *Lifetime Data Analysis* **15**:177–196.
- X. J. Hu, et al. (2003). ‘Regression parameter estimation from panel counts’. *Scandinavian Journal of Statistics* **30**:25–43.
- C. Y. Huang, et al. (2006). ‘Analysing panel count data with informative observation times’. *Biometrika* **93**:763–775.
- K. Y. Liang & S. L. Zeger (1986). ‘Longitudinal data analysis using generalized linear models’. *Biometrika* **73**:13–22.

- D. Y. Lin, et al. (2000). ‘Semiparametric regression for the mean and rate functions of recurrent events’. *Journal of the Royal Statistical Society, B Series* **62**:711–730.
- D. Y. Lin & Z. Ying (2001). ‘Semiparametric and nonparametric regression analysis of longitudinal data’. *Journal of the American Statistical Association* **96**:103–126.
- R. J. A. Little & D. B. Rubin (2002). *Statistical analysis with missing data*. Wiley Series in Probability and Statistics.
- R Development Core Team (2007). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-3.
- J. Sun & J. D. Kalbfleisch (1995). ‘Estimation of the mean function of point processes based on panel count data’. *Statistica Sinica* **5**:279–290.
- J. Sun, et al. (2007). ‘Regression analysis of panel count data with dependent observation times’. *Biometrics* **63**:1053–1059.
- J. Sun & L. J. Wei (2000). ‘Regression analysis of panel count data with covariate-dependent observation and censoring times’. *Journal of the Royal Statistical Society, Series B* **62**:293–302.
- J. A. Wellner & Y. Zhang (2000). ‘Two estimators of the mean of a counting process with panel count data’. *The Annals of Statistics* **28**:779–814.

J. A. Wellner & Y. Zhang (2007). ‘Two likelihood-based semiparametric estimation methods for panel count data with covariates’. *The Annals of Statistics* **35**:2106–2142.

Y. Zhang (2002). ‘A semiparametric pseudolikelihood estimation method for panel count data’. *Biometrika* **89**:39–48.

## Appendix

We assume that  $(N_i(\cdot), X_i, Z_i(\cdot), \tilde{N}_i^*(\cdot), \xi_i(\cdot))$  are *i.i.d.* quintuples of random processes over time  $t \in [0, \tau]$  for individuals 1 through  $n$ . We assume that the  $Z_i(\cdot)$  have bounded variations, that is  $|Z_{ji}(0)| + \int_0^\tau |dZ_{ji}(t)| \leq K$ ,  $j = 1, \dots, g$  and that the number of observations per person  $K_i$  is bounded. The inverse weight  $\rho_i(t; \gamma, \delta)$  must be finite and bounded away from zero, that is for all  $t \in [0, \tau]$   $\rho_i(t; \gamma_0, \delta) > c$  for some  $c > 0$ .

At point  $\{\beta_0, \gamma_0, \mathcal{A}_0\}$  the process  $\mathcal{M}_i(t)$  has zero mean for any fixed  $\delta$ . To see that, we proceed by showing that

$$E \left\{ \frac{N_i(s) d\tilde{N}_i(s)}{\rho_i(s; \gamma_0, \delta)} \middle| X_i \right\} = \xi_i(s) \exp\{(\beta_0 + \delta)^T X_i\} d\mathcal{A}_0(s),$$

using the iterated expectation formula by further conditioning on  $\{Z_i(t), X_i, N_i(t), C_i \geq t\}$  and using both assumptions of non-informative censoring for the outcome model (3) and the assumption on observation times (4). At time  $s$  such that



$C_i \geq s$

$$\begin{aligned}
 & E \left\{ E \left[ N_i(s) \frac{d\tilde{N}_i(s)}{\exp\{\gamma_0^T Z_i(s)\}} \middle| Z_i(s), X_i, N_i(s), C_i \geq s \right] \middle| X_i \right\} \\
 = & E \left\{ N_i(s) \frac{\exp\{\delta^T X_i\}}{\exp\{\gamma_0^T Z_i(s)\}} E \left[ d\tilde{N}_i^*(s) \middle| Z_i(s), X_i, N_i(s), C_i \geq s \right] \middle| X_i \right\} \\
 = & E \left\{ N_i(s) \frac{\exp\{\delta^T X_i\}}{\exp\{\gamma_0^T Z_i(s)\}} \exp\{\gamma_0^T Z_i(s)\} d\tilde{\Lambda}_0(s) \middle| X_i \right\} \\
 = & \exp\{\delta^T X_i\} d\tilde{\Lambda}_0(s) E \{N_i(s) | X_i\} \\
 = & \exp\{\delta^T X_i\} d\tilde{\Lambda}_0(s) \exp\{\beta_0^T X_i\} \Lambda_0(s) = \exp\{(\beta_0 + \delta)^T X_i\} d\mathcal{A}_0(s).
 \end{aligned}$$

When  $C_i < s$  the individual has been censored by time  $s$  and thus  $dN_i(s) = 0$ .

We outline here the large sample theory for the IIRR weighted approach. The large sample distribution of  $n^{1/2}(\hat{\beta}(\hat{\gamma}, \delta) - \beta_0)$  and the covariance matrix follow from the asymptotic normality of  $U(\beta_0; \gamma_0, \delta)$  together with a series of Taylor expansions. The consistency of  $\hat{\beta}$  follows from the facts that  $EU(\beta_0; \gamma_0, \delta) = 0$  under models (1) and (2) and that the estimator of  $\gamma_0$  is consistent.

The estimating function (12) at  $\beta_0$  can be written as

$$U(\beta_0; \gamma_0, \delta) = \sum_{i=1}^n \int_0^\tau w(t) [X_i - Av_2(X)(t; \beta_0, \delta)] d\mathcal{M}_i(t; \beta_0, \gamma_0, \delta, \mathcal{A}_0).$$

Further  $n^{-1/2}U(\beta_0; \gamma_0, \delta)$  is asymptotically equivalent to

$$\frac{1}{\sqrt{n}} \int_0^\tau w(s) \sum_{i=1}^n [X_i(s) - av_2(x)(s; \gamma_0, \delta)] \left[ \frac{N_i(s) d\tilde{N}_i(s)}{\rho_i(s; \gamma_0, \delta)} - \exp\{(\beta_0 + \delta)^T X_i\} \xi_i(s) d\mathcal{A}_0(s) \right].$$

A sequence of Taylor series expansions yield

$$\begin{aligned}
 \frac{1}{\sqrt{n}} U(\beta_0; \hat{\gamma}, \delta) &= \frac{1}{\sqrt{n}} U(\beta_0; \gamma_0, \delta) \\
 &\quad - \frac{1}{n} \frac{\partial U(\beta_0; \gamma, \delta)}{\partial \gamma} \Big|_{\gamma_0} \left( \frac{1}{n} \frac{\partial \tilde{U}(\gamma)}{\partial \gamma} \Big|_{\gamma^*} \right)^{-1} \frac{1}{\sqrt{n}} \tilde{U}(\gamma_0) \quad (17)
 \end{aligned}$$

with  $\gamma^\circ$  and  $\gamma^*$  being on the line segment between  $\gamma_0$  and  $\hat{\gamma}$ . Further, expansion of the estimating function (12) for the outcome model at  $(\hat{\beta}, \hat{\gamma}, \delta)$  around  $(\beta_0, \hat{\gamma}, \delta)$  is

$$U(\hat{\beta}; \hat{\gamma}, \delta) = U(\beta_0; \hat{\gamma}, \delta) + \frac{\partial U^T(\beta; \hat{\gamma}, \delta)}{\partial \beta} \Big|_{\beta^*} (\hat{\beta} - \beta_0),$$

with  $\beta^*$  being on the line segment between  $\hat{\beta}$  and  $\beta_0$ .

Let  $D$  and  $H$  be a matrix of derivatives of the estimating function  $U$  with respect to the recurrent events model parameter  $\beta$ , respectively the observation time model parameter  $\gamma$ , at its true value

$$\begin{aligned} D &\equiv \lim_{n \rightarrow \infty} E \left[ -\frac{1}{n} \frac{\partial U^T(\beta; \gamma_0, \delta)}{\partial \beta} \Big|_{\beta_0} \right] \\ &= E \int_0^\tau \frac{w(t)}{\rho_1(t; \gamma_0, \delta)} [X_1 - av_2(X)(t; \beta_0, \delta)]^{\otimes 2} \xi_1(t) \exp\{(\beta_0 + \delta)^T X_1\} d\mathcal{A}_0(t) \\ H &\equiv \lim_{n \rightarrow \infty} E \left[ -\frac{1}{n} \frac{\partial U^T(\beta_0; \gamma, \delta)}{\partial \gamma} \Big|_{\gamma_0} \right] \\ &= E \int_0^\tau -\frac{w(t)}{\rho_1(t; \gamma_0, \delta)} [X_1 - av_2(X)(t; \beta_0, \delta)] Z_1(t)^T N_1(t) d\tilde{N}_1(t). \end{aligned}$$

Define the covariance matrix  $V$  as

$$\begin{aligned} V &\equiv \lim_{n \rightarrow \infty} \text{Cov} \left[ \frac{1}{\sqrt{n}} U(\beta_0; \hat{\gamma}, \delta) \right] \\ &= \lim_{n \rightarrow \infty} \text{Cov} \left[ \frac{1}{\sqrt{n}} U(\beta_0; \gamma_0, \delta) - \frac{1}{n} \frac{\partial U^T(\beta_0; \gamma, \delta)}{\partial \gamma} \Big|_{\gamma_0} \left( \frac{1}{n} \frac{\partial \tilde{U}^T(\gamma)}{\partial \gamma} \Big|_{\gamma_0} \right)^{-1} \frac{1}{\sqrt{n}} \tilde{U}(\gamma_0) \right], \end{aligned}$$

where it is accounted for estimation of  $\gamma_0$  by including the second term on

right-hand side in  $V$ . Finally,

$$\begin{aligned} \sqrt{n}(\hat{\beta}(\hat{\gamma}, \delta) - \beta_0) &= \left( -\frac{1}{n} \frac{\partial U^T(\beta; \hat{\gamma}, \delta)}{\partial \beta} \Big|_{\beta^*} \right)^{-1} \frac{1}{\sqrt{n}} U(\beta_0; \hat{\gamma}, \delta) \\ &= \left( -\frac{1}{n} \frac{\partial U^T(\beta; \hat{\gamma}, \delta)}{\partial \beta} \Big|_{\beta^*} \right)^{-1} \left[ \frac{1}{\sqrt{n}} U(\beta_0; \gamma_0, \delta) \right. \\ &\quad \left. - \frac{1}{n} \frac{\partial U^T(\beta_0; \gamma, \delta)}{\partial \gamma} \Big|_{\gamma^0} \left( \frac{1}{n} \frac{\partial U^T(\gamma)}{\partial \gamma} \Big|_{\gamma^*} \right)^{-1} \frac{1}{\sqrt{n}} \tilde{U}(\gamma_0) \right] \end{aligned}$$

and thus  $\sqrt{n}(\hat{\beta}(\hat{\gamma}, \delta) - \beta_0)$  is asymptotically equivalent to

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n D^{-1} \left\{ \int_0^\tau w(t) [X_i(t) - av_2(x)(t; \beta_0, \delta)] d\mathcal{M}_i(t; \beta_0, \gamma_0, \delta, \mathcal{A}_0) \right. \\ &\quad \left. - HA^{-1} \int_0^\tau [Z_i(t) - av_1(z)(t; \gamma_0)]^T d\tilde{\mathcal{M}}_i(t; \gamma_0, \Lambda_0) \right\}. \end{aligned}$$

It is a sum of  $n$  mean zero i.i.d. random vectors for any  $\delta$  and thus  $\sqrt{n}(\hat{\beta}(\hat{\gamma}, \delta) - \beta_0)$  is zero mean Normal random vector with covariance matrix  $D^{-1}VD^{-1}$ . The consistency of  $V$  follows directly from arguments of Lin & Ying (2001) and therefore  $D^{-1}VD^{-1}$  is consistently estimated by  $\hat{D}^{-1}\hat{V}\hat{D}^{-1}$ . Using arguments similar to Liang & Zeger (1986), asymptotically equivalent expressions are obtained when using a random variable  $\hat{\delta}$  instead of arbitrary fixed  $\delta$ .

Table 1: Summary statistics for the simulation studies.

n	Estimator	$\gamma_1$	$\gamma_2$	Bias	SSE	BSE	CP	MSE
50	Independent	0	0	-0.01	0.33	0.32	0.96	0.11
		-0.6	0	-0.45	0.33	0.32	0.65	0.30
		-0.6	0.8	-1.17	0.32	0.31	0.07	1.46
	Dependent	0	0	-0.01	0.29	0.31	0.95	0.08
		-0.6	0	0.01	0.29	0.32	0.94	0.08
		-0.6	0.8	-0.52	0.27	0.28	0.74	0.47
	IIRR weighted	0	0	-0.00	0.29	0.31	0.96	0.08
		-0.6	0	0.00	0.29	0.32	0.94	0.08
		-0.6	0.8	0.04	0.30	0.32	0.95	0.09
100	Independent	0	0	-0.01	0.20	0.22	0.95	0.04
		-0.6	0	-0.45	0.24	0.23	0.49	0.25
		-0.6	0.8	-1.12	0.20	0.20	0.10	1.30
	Dependent	0	0	-0.01	0.17	0.22	0.95	0.03
		-0.6	0	0.00	0.22	0.23	0.96	0.05
		-0.6	0.8	-0.41	0.20	0.20	0.62	0.55
	IIRR weighted	0	0	-0.01	0.17	0.22	0.96	0.03
		-0.6	0	0.00	0.22	0.23	0.96	0.05
		-0.6	0.8	0.05	0.22	0.23	0.96	0.05

NOTE: Bias is the sampling mean of  $\hat{\beta}$  minus  $\beta_{01}$ , SSE is the sampling standard error of  $\hat{\beta}$ , BSE is the sampling mean of bootstrapped standard errors, CP is the 95% sampling coverage probability of the true  $\beta_{01}$  and MSE is the sampling mean squared error. Simulations are based on 1000 replicates.

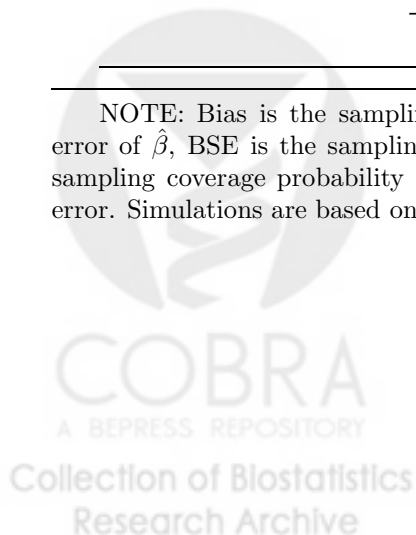


Table 2: Summary statistics for the simulation studies.

n	Estimator	$\gamma_1$	$\gamma_2$	Bias	SSE	BSE	CP	MSE
200	Independent	0	0	0.01	0.15	0.15	0.97	0.02
		-0.6	0	-0.43	0.15	0.16	0.28	0.21
		-0.6	0.8	-1.10	0.13	0.14	0.00	1.34
	Dependent	0	0	0.00	0.13	0.15	0.96	0.02
		-0.6	0	0.01	0.13	0.16	0.95	0.02
		-0.6	0.8	-0.62	0.15	0.14	0.43	0.52
	IIRR weighted	0	0	0.00	0.13	0.15	0.95	0.02
		-0.6	0	0.01	0.13	0.16	0.96	0.02
		-0.6	0.8	0.04	0.15	0.16	0.95	0.02
500	Independent	0	0	-0.01	0.09	0.10	0.96	0.01
		-0.6	0	-0.43	0.10	0.10	0.01	0.20
		-0.6	0.8	-1.12	0.09	0.09	0.00	1.27
	Dependent	0	0	-0.01	0.08	0.10	0.96	0.01
		-0.6	0	0.00	0.09	0.10	0.96	0.01
		-0.6	0.8	-0.57	0.09	0.09	0.26	0.41
	IIRR weighted	0	0	-0.01	0.08	0.10	0.97	0.01
		-0.6	0	0.00	0.09	0.10	0.95	0.01
		-0.6	0.8	0.04	0.09	0.10	0.95	0.01



Table 3: Summary statistics for the bladder cancer dataset.

Variable	Estimator	$\hat{\beta}$	SE( $\hat{\beta}$ )	p-val	95%CI
$X_1$	Independent	-1.297	0.409	<0.01	(-1.969,-0.625)
	Dependent	-1.491	0.430	<0.01	(-2.198,-0.784)
	IIRR weighted	-1.750	0.443	<0.01	(-2.479,-1.020)
$X_2$	Independent	0.236	0.078	<0.01	(0.107,0.364)
	Dependent	0.241	0.083	<0.01	(0.104,0.378)
	IIRR weighted	0.241	0.083	<0.01	(0.105,0.377)
$X_3$	Independent	-0.053	0.134	0.69	(-0.273,0.167)
	Dependent	-0.101	0.150	0.50	(-0.348,0.145)
	IIRR weighted	-0.098	0.151	0.52	(-0.346,0.151)

