



UW Biostatistics Working Paper Series

9-26-2005

Longitudinal Data Analysis for Generalized Linear Models under Irregular, Biased Sampling: Situations with Follow-up Dependent on Outcome or Auxiliary Outcome-related Variables

Petra Buzkova

University of North Carolina, buzkova@u.washington.edu

Thomas Lumley

University of Washington, tlumley@u.washington.edu

Suggested Citation

Buzkova, Petra and Lumley, Thomas, "Longitudinal Data Analysis for Generalized Linear Models under Irregular, Biased Sampling: Situations with Follow-up Dependent on Outcome or Auxiliary Outcome-related Variables" (September 2005). *UW Biostatistics Working Paper Series*. Working Paper 262.

<http://biostats.bepress.com/uwbiostat/paper262>

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

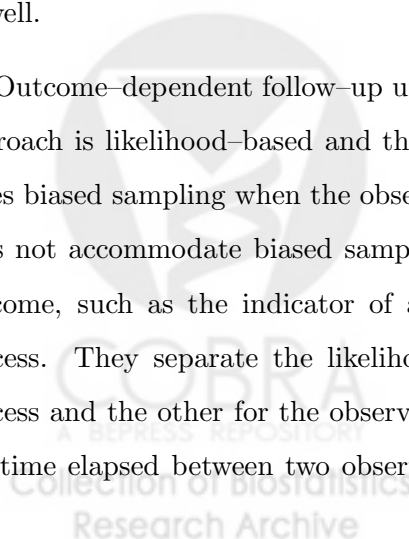
1. Introduction

Longitudinal data arise in many fields of biomedical research. In a longitudinal study, of focus is to examine the effect of covariate process $\{X_i(t), t \in [0, \tau]\}$ on the outcome process $\{Y_i(t), t \in [0, \tau]\}$, where the predetermined constant τ is the end of study or time at which the last person to follow drops out. Longitudinal data are often unbalanced and irregularly spaced resulting in a follow-up that is challenging for the statistical analysis of such data set. That occurs when designed sampling times are not strictly followed or when the sampling times are observational times with no direct control over the timing. An example of the first follow-up category, where there is an extra noise to the pre-defined observation times, is the health services research study data we analyze in Section 4. It is a study on a population of homeless people, where the investigators experienced an extreme non-compliance to the scheduled visits. An example of the second follow-up category are administrative data, where the actual sampling process is not part of the study. In both situations, the investigator has truly no control over the observation times and the related frequency of sampling the individuals. These irregular observation times can be specific to subject's characteristics. When the outcome and sampling are correlated conditional on the covariates of the outcome model then that introduces a biased sampling design for the outcome model. We demonstrate the philosophy of biased sampling with a simple example taken from air pollution. Assume an air pollution measure at time t as a covariate $X(t)$. Let the outcome $Y(t)$ be a lung function measure, such as FEV1, the volume exhaled during the first second of a forced expiratory maneuver started from the level of total lung capacity. Scientists are interested in quantifying the association of FEV and air pollution. Further assume a binary indicator, $Z(t)$, of an asthma attack at time t . The lung function measure clearly is associated both with the air pollution measure and presence or absence of an asthma attack. Also, the occurrence of an asthma attack may be related to the the air pollution measure. Assume that a person with an asthma attack searches for medical help more often and that the person has a lower lung function measure. So, data on individuals with present asthma attacks form the majority of the observed data. Then, when modeling FEV with the air pollution covariate we obtain unbiased estimates for those who have come for a visit, primarily people who suffer from an asthma attack on that day. However, we obtain an exaggerated estimate of the association of the lung function measure and the air pollution measure for the general public.

Under discrete time, when the observation times come from a finite set of points, the sampling issue can be viewed as a missingness problem. Taxonomy of missingness, as formalized by Rubin (1976), is based on factors that drive the seeing of observations. The key issue is whether the fact that data are missing is related to the values of the variables in the data set. For a complete survey of the current methodology of missing data we refer a keen reader to Little & Rubin (2002). In discrete time models we can view biased sampling as being equivalent to missingness at random (MAR) of the outcome given covariates X and Z , that is covariates of the outcome model as well as covariates of the observation-times model. It is informative missingness (IM) of the outcome given the covariates X of the outcome model. Other terms that are being used to describe biased sampling under discrete time are informative inter-mitten missingness or nonignorable non-monotone follow-up. Under continuous time, when the observation times come from an interval such as $[0, \tau]$, we can not talk about missingness in the original meaning. There the data are missing with probability one as the outcome is observed at discrete time points, not continuously over time as a curve.

Most statistical literature on longitudinal data with unbalanced or irregular follow-up focuses on drop out, also called right censoring, under discrete observation times. Drop-out under continuous observation times is presented by Wu & Carroll (1988), Diggle & Kenward (1994) and Scharfstein et al. (1999). A few articles deal with intermittent missingness under discrete times, among them the pseudolikelihood procedures of Troxel et al. (1998) suitable for continuous outcome, the autoregressive approach of Albert et al. (2002) for binary outcome and the weighted generalized estimating equations approach of Preisser et al. (2000) for binary outcome as well.

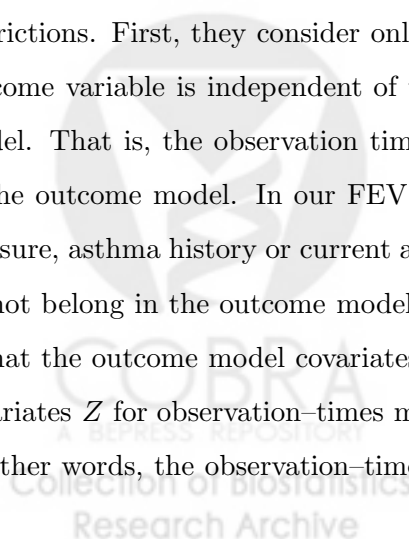
Outcome-dependent follow-up under continuous time addresses Lipsitz et al. (2002). Their approach is likelihood-based and thus, through the nature of likelihood method, it accommodates biased sampling when the observation times depend on the previous value of outcome. It does not accommodate biased sampling when there are additional covariates, associated with outcome, such as the indicator of asthma in our example above, that govern the sampling process. They separate the likelihood function into two components: one for the outcome process and the other for the observation-times process, where they compute the likelihood of the time elapsed between two observation times. In order to ignore the later process, Lipsitz



et al. impose a strong observation–times model assumption of dependence only on history of observed outcome measurements. Their model is limited to a linear regression model where the repeated measures of outcome have a multivariate Gaussian distribution. Moreover, their procedure relies for consistency of the estimators of parameters of interest on correct specification of the autocorrelation structure of outcome. Lipsitz et al. themselves point out that “the potential bias due to misspecification of the covariance can be considerable”.

Recently, H. Lin et al. (2004) developed a class of inverse–intensity of visit process–weighted estimators, a solution to an estimating equation. The outcome might be observed in a continuous-time fashion and the weighting, similar to ours, in their estimating equation accommodates biased sampling. However, their outcome model covariates are fixed over time. Next, their observation–times model does not allow for pre–specified visit times. Also, most important, their estimator is fairly complicated, involving specific smoothing techniques of baseline intensity of visiting to achieve \sqrt{n} consistency of the estimator.

We adopt approach of D. Lin & Ying (2001) integrating counting processes techniques into analysis of longitudinal data. Their estimation approach is based upon a set of estimating equations for both the observation–times model and the outcome model. Their estimator resembles a weighted least squares estimator. They use fully marginal models where the assumption $E[Y_i(t)|\{X_i(s), s \in [0, \tau]\}] = E[Y_i(t)|X_i(t)]$ is not needed, as discussed in Pepe & Anderson (1994). This assumption can be hard to satisfy. History of covariates can be incorporated into the current covariates. But supposing that conditioning on future covariates on top of the current one does not alter the mean model is a strong assumption. However, they impose a few restrictions. First, they consider only a linear regression model. Second, they assume that the outcome variable is independent of the observation times given the covariates of the outcome model. That is, the observation times are not allowed to depend on additional covariates not in the outcome model. In our FEV example the visit timing might depend on previous FEV measure, asthma history or current asthma severity or other aspects of the patient’s health that do not belong in the outcome model as covariates. Our philosophy behind choosing covariates is that the outcome model covariates X should be picked on scientific ground purpose but the covariates Z for observation–times model should cover the true nature of the sampling process. In other words, the observation–times model should be predictive whereas the outcome model



is an association model where estimation of contrast parameters is of major scientific interest.

In this paper we address parameter estimation under continuous time under generalized linear models, naturally handling continuous, binary and count outcomes, where the follow-up is informative for outcome given the covariates of the outcome model. We call those estimators “inverse-intensity rate-ratio-weighted” (IIRR) estimators. Standard statistical software, e.g. R Development Core Team (2004), can be used to compute the IIRR estimates.

Notation that we use and the class of models for outcome and the observation-times model is described in Section 2. Our estimation approach is given in Section 3. We illustrate our methods on a health service research study in Section 4. Homeless people with mental illness were randomized to three different treatments. Percentage days homeless within the last three months as an outcome variable and a handful of covariates were recorded at follow-up times not fixed by design. Treatment efficacy is the scientific question of our interest. This data set was used in H. Lin et al. (2004). We report results from a simulation study to demonstrate the estimators performance under finite sample size and to compare the squared errors under the independent GEE with probit/logit link estimators in Section 5.

2. Notation and models

We consider a fully marginal mean model for outcome Y as a function of covariates X of individual $i \in \{1, \dots, n\}$ at time $t \in [0, \tau]$. Thus we model $E[Y_i(t)|X_i(t)]$, denoted by $\mu_i(t)$ and called “the outcome model”. We focus on a class of generalized linear regression models, discussed in McCullagh & Nelder (1989)

$$g(\mu_i(t)) = \beta_0^T X_i(t). \quad (1)$$

The known link function g is monotonic and differentiable. Variance of outcome is $v(\mu_i(t))\phi$ with v a known positive function of the mean and ϕ a positive scale parameter. The outcome model (1) may be either based on a likelihood theory for exponential family distributions or a quasi-likelihood approach. The effect of time-varying covariates X_i is, through the link function g , modeled parametrically in a linear way. The parameter of major interest, β_0 , is a p dimensional vector.

The model for outcome, formulated in equation (1), is a functional full data model. We

do not, however, observe the outcome continuously over time but at certain times only, that we call sampling times or observation times. Denote for individual $i \in \{1, \dots, n\}$ the set of observation times $\{T_{i1}, T_{i2}, \dots, T_{iK_i}\}$ as \mathcal{T}_i , with $0 \leq T_{i1} < T_{i2} < \dots < T_{iK_i} \leq \tau$. For the i -th individual, K_i is a random total number of observed events. Denote $\mathbb{T} = \{\mathcal{T}_j, j = 1, \dots, n\}$ the set of sample's observation times. Define $N_i(t) = \sum_{k=1}^{K_i} I(T_{ik} \leq t)$ the counting process of number of events of individual i by time t . The underlying uncensored process we denote as $N_i^*(\cdot)$, $N_i(t) = N_i^*(t \wedge C_i)$ where C_i is drop-out time or end of follow-up τ , whatever comes first.

For each individual $i \in \{1, \dots, n\}$ at time $t \in [0, \tau]$ we adopt a marginal rate model for uncensored observation times that we call an observation-times model or a sampling-times model:

$$E[dN_i^*(t)|Z_i(t)] = \exp\{\gamma_0^T Z_i(t)\}d\Lambda_0(t). \quad (2)$$

Here, $\Lambda_0(\cdot)$ is an unspecified baseline cumulative (thus non-decreasing function of time t) hazard function, whereas γ_0 is unknown regression parameter. The $\Lambda_0(\cdot)$ is very flexible as we require it to be only continuous up to countably many points. In our practical settings a finite number of points suffices.

Two crucial assumptions are non-informative drop-out for the mean of outcome,

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

saying that $EY_i(t)$ depends on covariates $X_i(t)$ and drop-out C_i through covariates $X_i(t)$ only, and independent sampling assumption,

$$E[dN_i^*(t)|Z_i(t), X_i(t), Y_i(t), C_i \geq t] = E[dN_i^*(t)|Z_i(t)], \quad (4)$$

saying that observation times depend on covariates Z_i, X_i , on outcome Y_i and drop-out C_i through covariates Z_i only.

D. Lin & Ying (2001) require in their estimation approach an assumption that does exclude biased sampling. It is that the outcome variable is assumed independent of the observation times given the covariates of the outcome model. That is, the observation times are not allowed to depend on additional auxiliary variables not in the outcome model; covariates Z must be part of covariates X . In our example of studying an association between a lung function measure and the air pollution we would either need to include adjusting for the asthma attack in the

outcome model or assume an untrue assumption of independence of the sampling–times process and the lung function measure conditioning on the air pollution alone. Lin & Ying’s independent sampling assumption is

$$E[dN_i^*(t)|Z_i(t), X_i(t), Y_i(t), C_i \geq t] = E[dN_i^*(t)|X_i(t)]. \quad (5)$$

The difference between our independent sampling assumption (4) and their independent sampling assumption (5) is conditioning on covariates Z instead of X on the right hand side of the equations. However, when we allow the two sets of covariates to be arbitrary, we can introduce a biased sampling scheme into our outcome model and thus to obtain consistent estimates of β_0 we need to account for the biased sampling.

We write $\xi_i(t) = I(C_i > t)$ for the at-risk process and assume that $\Pr(C_i \geq \tau) > 0$. Additional technical assumptions are given in the Appendix A.

Note, that although outcome $Y_i(\cdot)$ is observed only at random times T_{ij} , the expectation in (1) does not condition on these times. Similar to Lipsitz et al. (2002), D. Lin & Ying (2001) and H. Lin et al. (2004), it is a functional or process–like approach. We model the conditional mean of outcome for a certain individual for any time t . Unlike for instance X. Lin & Carroll (2001), we do not condition on the set of the individual’s observation times. We say that at any time t the conditional mean outcome exists and follows our given model or that it only exists and follows the given model at the observation times but if we had observed outcome at a different time it would have existed and followed the given model.

3. Estimation

3.1. Biased sampling

We use an approach based on inverse–intensity–rate–ratio weighting. For individual $i \in \{1, \dots, n\}$ at time $t \in [0, \tau]$ define inverse weights $\rho_i(t; \gamma, h)$ as

$$\rho_i(t; \gamma, h) = \frac{\exp\{\gamma^T Z_i(t)\}}{h(X_i(t))}. \quad (6)$$

In the denominator we can include any function $h(\cdot)$ that is a deterministic function of the outcome model covariates $X_i(t)$. The inverse weight is proportional to the probability of individual i , relative to other individuals, having an observation at time t under the observation–times

model (2). Motivated by Hernán et al. (2002) we find it useful to incorporate a function $h_0(\cdot)$ that decreases the variability of the weights. We pick

$$h_0(X_i(t)) = \exp\{\delta_0^T X_i(t)\}$$

and we call the weight $\rho_i(t; \gamma, h_0)$ a stabilizing weight. We find the estimator of δ_0 when fitting a proportional rate model, similar to model (2), conditioning on covariates X only. When observation–times model covariates Z are a subset of the outcome model covariates X , then the inverse weight $\rho_i(t; \gamma, h_0)$ equals one for all individuals at all times, using the independent sampling assumption (4).

3.2. Observation–times model

Based on the proportional rates model (2) and the drop-out part of assumption (4), a parameter vector γ_0 of length g can be consistently estimated by $\hat{\gamma}$, the solution to a set of estimating equations $U^\dagger(\hat{\gamma}) = 0$. The estimating function $U^\dagger(\gamma)$ is defined as

$$U^\dagger(\gamma) = \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(t; \gamma)\} dN_i(t), \quad (7)$$

where the weighted average for variable Z at time t is defined as

$$\bar{Z}(t; \gamma) = \sum_{i=1}^n Z_i(t) \frac{\xi_i(t) \exp\{\gamma^T Z_i(t)\}}{\sum_{j=1}^n \xi_j(t) \exp\{\gamma^T Z_j(t)\}}. \quad (8)$$

The weights there are proportional to the probability of the individual having an observation at time t . Estimation of γ_0 is β_0 -free. Solution of (7) and derivation of asymptotic properties of the estimator are based on a zero mean random process $\mathcal{M}_i(\cdot; \gamma_0, \Lambda_0(\cdot))$ defined as

$$\mathcal{M}_i(t; \gamma, \Lambda(\cdot)) = N_i(t) - \int_0^t \xi_i(s) \exp\{\gamma^T Z_i(s)\} d\Lambda(s). \quad (9)$$

Though the estimating function (7) is the same as under the Cox proportional hazards model, the asymptotic variance is different due to imposing weaker assumptions in the proportional rate model (2). Define the asymptotic weighted mean curve of a covariate process $\{Z(t), t \in [0, \tau]\}$ as

$$\bar{z}(t; \gamma) = \lim_{n \rightarrow \infty} \bar{Z}(t, \gamma) = \frac{E[Z_1(t) \xi_1(t) \exp\{\gamma^T Z_1(t)\}]}{E[\xi_1(t) \exp\{\gamma^T Z_1(t)\}]}.$$

The asymptotic variance of $\sqrt{n}(\hat{\gamma} - \gamma_0)$ is Γ , $\Gamma = A^{-1}\Sigma A^{-1}$, where

$$\begin{aligned} A &\equiv \lim_{n \rightarrow \infty} E \frac{1}{n} \left[\frac{-\partial U^\dagger(\gamma)}{\partial \gamma} \Big|_{\gamma_0} \right] \\ &= E \int_0^\tau [Z_1(t) - \bar{z}(t; \gamma_0)]^{\otimes 2} \xi_1(t) \exp\{\gamma_0^T Z_1(t)\} d\Lambda_0(t) \\ \Sigma &\equiv \lim_{n \rightarrow \infty} \text{Cov} \left[\frac{1}{\sqrt{n}} U^\dagger(\gamma_0) \right] \\ &= E \left[\int_0^\tau [Z_1(t) - \bar{z}(t; \gamma_0)] d\mathcal{M}_1(t; \gamma_0, \Lambda_0(\cdot)) \int_0^\tau [Z_1(u) - \bar{z}(u; \gamma_0)]^T d\mathcal{M}_1(u; \gamma_0, \Lambda_0(\cdot)) \right]. \end{aligned} \quad (10)$$

Notation $x^{\otimes 2} = xx^T$ stands for the outer product of a vector x . The matrix A is further used in the formula for variance of estimator of β_0 in the outcome model (1) that is dependent upon estimation of the parameter γ_0 . There is a straightforward consistent estimator of Γ denoted by $\hat{\Gamma}$, $\hat{\Gamma} = \hat{A}^{-1}\hat{\Sigma}\hat{A}^{-1}$, where

$$\begin{aligned} \hat{\Sigma} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau [Z_i(t) - \bar{Z}(t; \hat{\gamma})] d\widehat{\mathcal{M}}_i(t) \sum_{k=1}^n \int_0^\tau [Z_k(u) - \bar{Z}(u; \hat{\gamma})]^T d\widehat{\mathcal{M}}_k(u) \\ \hat{A} &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau [Z_i(t) - \bar{Z}(t; \hat{\gamma})]^{\otimes 2} \xi_i(t) \exp\{\hat{\gamma}^T Z_i(t)\} d\widehat{\Lambda}(t) \end{aligned}$$

with

$$d\widehat{\mathcal{M}}_i(t) = N_i(t) - \int_0^t \xi_i(s) \exp\{\hat{\gamma}^T Z_i(s)\} d\widehat{\Lambda}(s)$$

and Aalen–Breslow estimator of $\Lambda_0(t)$

$$\widehat{\Lambda}(t) = \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{\sum_{j=1}^n \xi_j(s) \exp\{\hat{\gamma}^T Z_j(s)\}}.$$

See D. Lin et al. (2000) for detailed derivation of the parameter estimation in the observation–times model (2) and comparison to the widely used proportional mean model.

3.3. Outcome model

Motivated by generalized estimating equations, GEE, we define our estimating function as

$$\begin{aligned} U(\beta; \gamma, h) &= \sum_{i=1}^n \int_0^\tau W(t) X_i(t) \left\{ \frac{dg(\mu)}{d\mu} \Big|_{\mu_i(t; \beta)} \right\}^{-1} v(\mu_i(t; \beta))^{-1} \times \\ &\quad \times [Y_i(t) - \mu_i(t; \beta)] \frac{1}{\rho_i(t; \gamma, h)} dN_i(t). \end{aligned} \quad (11)$$

The estimating equation is

$$U(\beta; \hat{\gamma}, h) = 0.$$

Estimator $\hat{\gamma}$ of the observation–times proportional rates model parameter γ_0 is based on setting the estimating function (7) equal to zero. The estimating function (11) resembles GEE under the assumption of working independence that is being weighted by a function of time $W(t)$ but mainly by the inverse weight $\rho_i(t; \gamma, h)$. The weight $W(t)$ can for instance reflect the reliability of the data over time. D. Lin & Ying (2001) discuss that topic on page 106.

When we use a canonical link for a given exponential family distribution, the estimating function (11) simplifies to

$$U(\beta; \gamma, h) = \sum_{i=1}^n \int_0^\tau W(t) X_i(t) [Y_i(t) - \mu_i(t; \beta)] \frac{1}{\rho_i(t; \gamma, h)} dN_i(t).$$

At point $\{\beta_0, \gamma_0\}$ equation (11) has zero mean for any function h of covariates $X_i(t)$. To see that, we proceed as follows:

$$\begin{aligned} EU(\beta_0; \gamma_0, h) &= E \sum_{i=1}^n \int_0^\tau W(t) X_i(t) \left\{ \frac{dg(\mu)}{d\mu} \Big|_{\mu_i(t; \beta_0)} \right\}^{-1} v(\mu_i(t; \beta_0))^{-1} \times \\ &\quad \times [Y_i(t) - \mu_i(t; \beta_0)] \frac{h(X_i(t))}{\exp\{\gamma_0^T Z_i(t)\}} dN_i(t) \\ &= E \sum_{i=1}^n \int_0^\tau W(t) X_i(t) \left\{ \frac{dg(\mu)}{d\mu} \Big|_{\mu_i(t; \beta_0)} \right\}^{-1} v(\mu_i(t; \beta_0))^{-1} \times \\ &\quad \times h(X_i(t)) E \left\{ [Y_i(t) - \mu_i(t; \beta_0)] \frac{dN_i(t)}{\exp\{\gamma_0^T Z_i(t)\}} \Big| X_i(t) \right\}. \end{aligned}$$

Using iterated expectation formula by further conditioning on $\{Z_i(t), Y_i(t) | X_i(t)\}$ and using both assumptions of non–informative drop–out for the outcome model as defined in equation (3) and the independent sampling assumption defined in equation (4) we conclude that

$$E \left\{ [Y_i(t) - \mu_i(t; \beta_0)] \frac{dN_i(t)}{\exp\{\gamma_0^T Z_i(t)\}} \Big| X_i(t) \right\} = 0$$

for any $i \in \{1, \dots, n\}$ and any time $t \in [0, \tau]$.

Similarly to the generalized estimating equations, the estimator of β_0 can be obtained in an iterative procedure based on linearization in β of the estimating function (11). Solution to $U(\beta; \hat{\gamma}, h) = 0$ is \sqrt{n} -consistent and asymptotically normal. The asymptotic variance of $\sqrt{n}(\hat{\beta}(\hat{\gamma}, h) - \beta_0)$ is

$$D^{-1}VD^{-1}.$$

The square matrix of derivatives D is defined as

$$D \equiv \lim_{n \rightarrow \infty} E \left[-\frac{1}{n} \frac{\partial U(\beta; \gamma_0, h)}{\partial \beta} \Big|_{\beta_0} \right]$$

and the square matrix of covariances V as

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

We account for estimation of γ_0 by including the second term on right-hand side. The asymptotic variance is consistently estimated by $\hat{D}^{-1} \hat{V} \hat{D}^{-1}$, where we replace the true value of the parameters with their estimates in the matrix D . In canonical links, this simplifies to

$$\hat{D} = \frac{1}{n} \sum_{i=1}^n \int_0^\tau W(t) X_i(t) \frac{\partial \mu_i(t; \beta)}{\partial \beta} \Big|_{\hat{\beta}} \frac{1}{\rho_i(t; \hat{\gamma}, h)} dN_i(t).$$

A consistent estimator of the matrix V is

$$\begin{aligned} \hat{V} &= \frac{1}{n} \sum_{i=1}^n \left[\int_0^\tau W(t) X_i(t) \left\{ \frac{dg(\mu)}{d\mu} \Big|_{\mu_i(t; \hat{\beta})} \right\}^{-1} v(\mu_i(t; \hat{\beta}))^{-1} \times \right. \\ &\quad \times \left[Y_i(t) - \mu_i(t; \hat{\beta}) \right] \frac{1}{\rho_i(t; \hat{\gamma}, h)} dN_i(t) - \\ &\quad \left. - \hat{H} \hat{A}^{-1} \int_0^\tau [Z_i(t) - \bar{Z}(t; \hat{\gamma})]^T d\mathcal{M}_i(t, \hat{\gamma}, \hat{\Lambda}(\cdot)) \right]^{\otimes 2}, \end{aligned}$$

where the random process $\{\mathcal{M}(t), t \in [0, \tau]\}$ is defined in equation (9), the matrix \hat{A} in (11) and \hat{H} is

$$\begin{aligned} \hat{H} &= - \sum_{i=1}^n \int_0^\tau W(t) X_i(t) \left\{ \frac{dg(\mu)}{d\mu} \Big|_{\mu_i(t; \hat{\beta})} \right\}^{-1} v(\mu_i(t; \hat{\beta}))^{-1} \times \\ &\quad \times \left[Y_i(t) - \mu_i(t; \hat{\beta}) \right] \frac{1}{\rho_i(t; \hat{\gamma}, h)} Z_i(t)^T dN_i(t). \end{aligned}$$

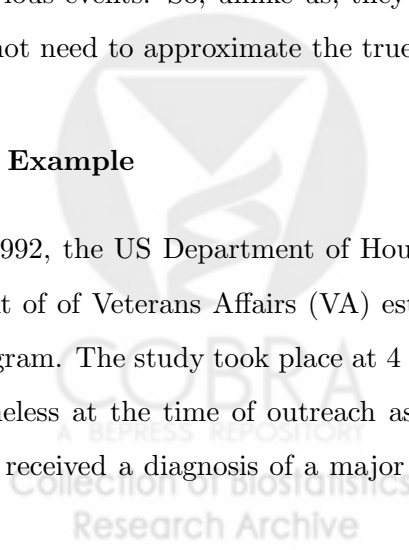
Instead of using a general function h for optimality reasons we prefer using the $h_0(\delta_0, X(t))$ function as explained in Section 3.1. As δ_0 is unknown we estimate it. We have derived the asymptotic properties generally for any h and thus for any δ fixed. But using arguments similar to Liang & Zeger (1986), asymptotically we obtain equivalent expressions when plugging in a random variable $\hat{\delta}$ instead of any fixed δ , for any $\hat{\delta}$ with asymptotic standard error $O_p(n^{1/2})$.

A recently published paper H. Lin et al. (2004) supports our claim that scenario studied in this paper is of a scientific interest. They approach the same problem as we do, namely

analysis of longitudinal data under continuous times with irregular, outcome-dependent follow-up and suggest a class of consistent and asymptotically normal estimators for the parameters of marginal generalized linear model. There are many similarities between their approach and ours and their solutions and ours. However, they do not enable to include time-varying covariates in the outcome model but instead assume fixed covariates. They assume that censoring is independent of outcome given outcome model covariates, a stronger assumption to our non-informative drop-out assumption (3) about mean of outcome. To model the sampling times they use a similar Cox-type intensity model with a similar assumption to our independent sampling assumption (4). Unfortunately, to estimate consistently the parameter of primary interest β_0 , they need an estimator of the baseline intensity to be consistent at a rate bigger than $n^{\frac{1}{4}}$. In order to do that they use a kernel-smoothed estimator with a sample-size-dependent bandwidth based on the Breslow's estimator of the cumulative baseline intensity. They start with a general sampling intensity model but then, due to the curse of dimensionality, assume a proportional hazard form. Their sampling-times model differs from ours in the sense that they require conditioning on the observed history whereas our model is a fully marginal model allowing for that but not requiring it. The sampling-times model, that we refer to as "without biased sampling" they call a "null" model. The sampling-times model we refer to as "under biased sampling" they call "predictor adjusted" model. They run into the same problem in the sampling-times model with observing covariate process over time. To avoid the necessity of knowledge of the process of sampling-times model covariates over time they base their sampling-times model directly on the covariates values observe at the previous event or all the history of previous events. So, unlike us, they assume a model that is unlikely to be true but then they do not need to approximate the true covariate process to arrive at a feasible estimator.

4. Example

In 1992, the US Department of Housing and Urban Development (HUD) and the US Department of Veterans Affairs (VA) established the HUD-VA Supported Housing (HUD-VASH) program. The study took place at 4 sites across the country. Veterans were eligible if they were homeless at the time of outreach assessment, had been homeless for 1 month or longer, and had received a diagnosis of a major psychiatric disorder or an alcohol or drug abuse disorder.



All veterans provided written informed consent to participate in the study. The 460 homeless veterans were randomly assigned to 1 of 3 intervention groups:

- HUD–VASH intervention consisting of case management and housing vouchers (182 individuals);
- case management (90 individuals);
- standard VA homeless services (188 individuals).

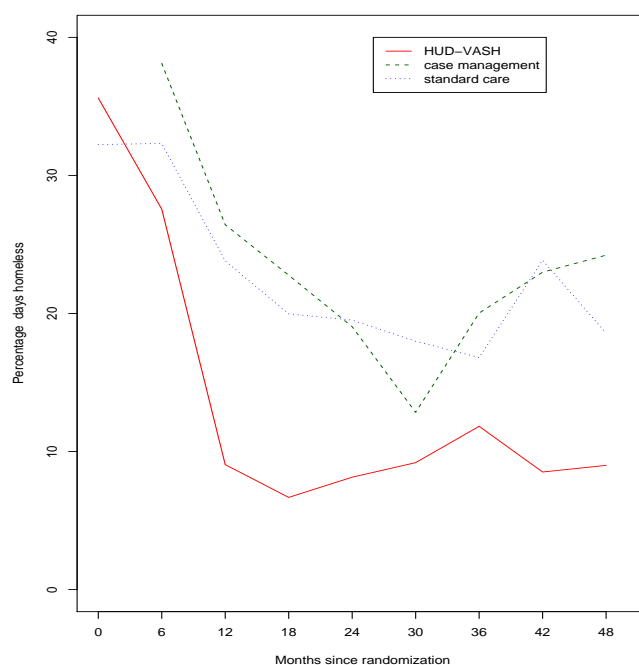
Vouchers authorized payment of a standardized local fair–market rent less 30% of the individual beneficiary’s income. The important question is whether setting aside housing resources is either necessary or sufficient for facilitating exit from homelessness in this population. The primary outcome was percentage of days homeless during the last 3 months. Auxiliary time–dependent variables collected during the study were income in the past three months and whether social security or VA benefits were received during the past three months. Follow–up interviews were scheduled for every 3 months. However, subjects often missed assessment and came between scheduled interviews. Concern is raised that there is an association between the visit process and the outcome process. For detailed study description see Rosenheck et al. (2003). It also contains cost-effectiveness considerations for the 3 interventions.

In the analysis of the data, we set τ to 48 months and $C_i = \tau$ for all individuals $i \in \{1, \dots, 460\}$. That means that we do not allow anybody to drop–out of the study sooner than at the 48 months. There is not any drop–out by protocol that would exclude certain individuals after study beginning and if no event occurs by the study end we consider that just an intermittent missing data. The 460 individuals made a total of 2855 follow–up visits by 48 months since randomization. Quantiles of the total counts of follow–up visits per treatment arm, shown in Table 1, suggest highest follow–up for the HUD–VASH intervention group, lower for the case–management group and lowest for the standard VA care. Figure 1 shows the average percentage homeless during the last 3 months specific for each treatment group. In both figures the time discretization is based on 6 months intervals. A crude view at the data suggests that the HUD–VASH intervention is more effective in reducing homelessness than the other two interventions that appear comparable. The HUD–VASH intervention group has the highest level of follow–up visits and the standard care group the lowest level of visiting.

Table 1: HUD–VASH: quantiles of number of follow–up visits per individual by treatment arm.

| | minimum | 25% | median | 75% | maximum |
|-----------------|---------|-----|--------|-----|---------|
| HUD–VASH | 1 | 7 | 9 | 10 | 12 |
| case management | 1 | 5 | 7 | 9 | 12 |
| standard care | 1 | 3 | 6 | 8 | 12 |

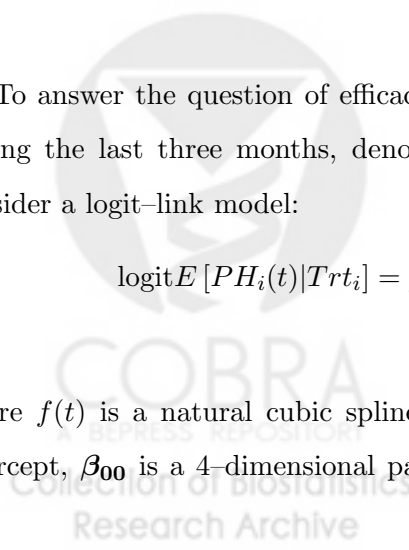
Figure 1: HUD–VASH: averaged percentage days homeless during the last three months by treatment arm.



To answer the question of efficacy of intervention we model the percentage days homeless during the last three months, denoted as PH , as a function of treatment assignment. We consider a logit–link model:

$$\begin{aligned} \text{logit}E [PH_i(t)|Trt_i] = & \beta_{00}f(t) + \beta_{01}I(Trt_i = \text{HUD–VASH}) \\ & + \beta_{02}I(Trt_i = \text{case management}), \end{aligned} \tag{12}$$

where $f(t)$ is a natural cubic spline with 4 degrees of freedom parametrically modeling the intercept, β_{00} is a 4–dimensional parameter vector and $\beta_0 = (\beta_{01}, \beta_{02})^T$ is the 2–dimensional



parameter vector of primary interest. It quantifies the association of the outcome and interventions contrasts. The standard VA homeless service arm is the reference group.

The sampling-times model we define in equation (13) as a proportional rate model. The covariates were suggested by the primary investigator Dr. Rosenheck. The time-invariant predictors of timing of visits are intervention assignment, income at baseline, in thousands of dollars, denoted as IB , an indicator of receiving any social security or VA benefits at baseline, BB and a Lehman measure of the quality of life at baseline. Time-varying predictors for the sampling-times model are percentage homeless approximated by previous value carried forward, denoted by PH^* , and cumulative number of visits so far, denoted by N_- .

$$\begin{aligned}
 E [dN_i^*(t) | Trt_i, IB_i, BB_i, PH_i^*(t), QLB_i, N_{-i}(t)] = & \exp \{ \gamma_{01} I(Trt_i = \text{HUD-VASH}) + \\
 & + \gamma_{02} I(Trt_i = \text{case management}) + \gamma_{03} IB_i + \gamma_{04} BB_i + \gamma_{05} PH_i^*(t) + \gamma_{06} QLB_i + \\
 & + \gamma_{07} N_{-i}(t) + \gamma_{08} N_-^{\text{HUD-VASH}}(t) + \gamma_{09} N_-^{\text{case management}}(t) \} d\Lambda_0(t)
 \end{aligned} \tag{13}$$

We estimate the parameter γ_0 of dimension 9. Parameter estimates, as shown in Table 2, suggest that higher intensity of visiting is associated with lower baseline income, lower baseline quality of life, receiving any social or VA benefits at baseline, having higher approximated percentage days homeless and higher cumulative number of visits so far, differentiated by treatment arm. At any time, individual in the HUD-VASH intervention arm is more likely to have a visit than an individual under only case management, comparing two individuals on the same level of baseline income, with the same quality of life baseline measure, indicator of social or VA benefits at baseline, approximated percentage days homeless and having the same number of visits so far, assuming that the number of visits so far is in the range of (0, 11). Similarly, individual under case management is more likely to have a visit than an individual on standard care, comparing two individuals on the same level of baseline income, with the same quality of life baseline measure, indicator of social or VA benefits at baseline, approximated percentage days homeless and having the same number of visits so far, ranging from 0 to 11.

There is positive statistically significant difference on the 5% level of proportion of days homeless within the last 3 months between the HUD-VASH intervention and standard VA care, favoring the HUD-VASH intervention. Comparing the two treatment arms in averaged

Table 2: HUD–VASH: parameter estimates and their standard errors in the intensity rate model (13) for sampling times.

| | $\hat{\gamma}_0$ | $SE(\hat{\gamma}_0)$ |
|--------------------------|------------------|----------------------|
| HUD–VASH | 0.359 | 0.044 |
| case management | 0.217 | 0.054 |
| IB | -0.172 | 0.482 |
| BB | 0.104 | 0.041 |
| PH^* | 0.001 | 0.001 |
| QLB | -0.007 | 0.019 |
| N_- | 0.044 | 0.015 |
| $N_-^{HUD-VASH}$ | -0.018 | 0.016 |
| $N_-^{case\ management}$ | -0.014 | 0.023 |

percentage days homeless during the last 3 months under logit model is time specific, but we can compare the two treatment arms by the odds ratio of probability of being homeless during the last 3 months. There, the odds ratio is 0.42, with 95% confidence interval (0.28, 0.62). The estimate of β_{02} suggests decrease of proportion days homeless comparing the case management group to the standard VA care group. However, we lacked power to show that the outcomes in the two arms really differ. These findings support the conclusion that setting aside housing resources through the HUD–VASH intervention is necessary and sufficient for facilitating exit from homelessness in this population.

Table 3: HUD–VASH: estimates of primary parameter of interest $\beta_0 = (\beta_{01}, \beta_{02})^T$ for model (12). Standard errors are included.

| | Intervention group | |
|---------------------|--------------------|-----------------|
| | HUD–VASH | Case management |
| $\hat{\beta}_0$ | -0.879 | -0.028 |
| $SE(\hat{\beta}_0)$ | 0.210 | 0.223 |

We also computed the naive estimates using GEE that do not adjust for the possibility of biased sampling. The naive parameter estimates, shown in Table 4, suggest qualitatively the same answer. However, we see a decrease in favoring the HUD–VASH treatment and the estimate of β_{02} has a positive sign. That suggests that the case management intervention has a negative effect which is however not significantly different from zero on 5% α level. Fitting the sampling–times model (13) we learned that individuals who were worse off, which is those with more homelessness, lower baseline income and receiving baseline benefits, tended to have increased intensity of visiting. Similar to H. Lin et al. (2004), we conclude that the data tend to be biased upwards. This biasness is deferent for the treatment arms, as suggested by fitting a sampling–times model with interaction of treatment arm and the other covariates.

Table 4: HUD–VASH: naive GEE estimates of primary parameter of interest $\beta_0 = (\beta_{01}, \beta_{02})^T$ for parametric model (12). Standard errors are included.

| | Intervention group | |
|---------------------|--------------------|-----------------|
| | HUD–VASH | Case management |
| $\hat{\beta}_0$ | -0.774 | 0.081 |
| $SE(\hat{\beta}_0)$ | 0.208 | 0.163 |

Our results support the idea that the bias due to informative follow–up can be avoided if the sampling times are modeled properly and the outcome model estimation takes that into consideration.

5. Simulation Study

We consider a logistic regression model

$$E[Y_i(t)|X_i(t)] = \frac{\exp\{\alpha_0(t) + \beta_0 X_i(t)\}}{1 + \exp\{\alpha_0(t) + \beta_0 X_i(t)\}}, \quad (14)$$

where the parameter of primary interest is β_0 of dimension 1 and we are interested in the association of mean of the outcome and the covariate X at time t . We simulate the data using a probit link, because under probit link marginalization is easy and logit and probit are very

similar functions, once normalized by $\pi/\sqrt{3}$, see Figure 2. That is

$$\Phi\left(\frac{\alpha_0(t) + \beta_0 X_i(t)}{\pi/\sqrt{3}}\right) \approx \frac{\exp\{\alpha_0(t) + \beta_0 X_i(t)\}}{1 + \exp\{\alpha_0(t) + \beta_0 X_i(t)\}}.$$

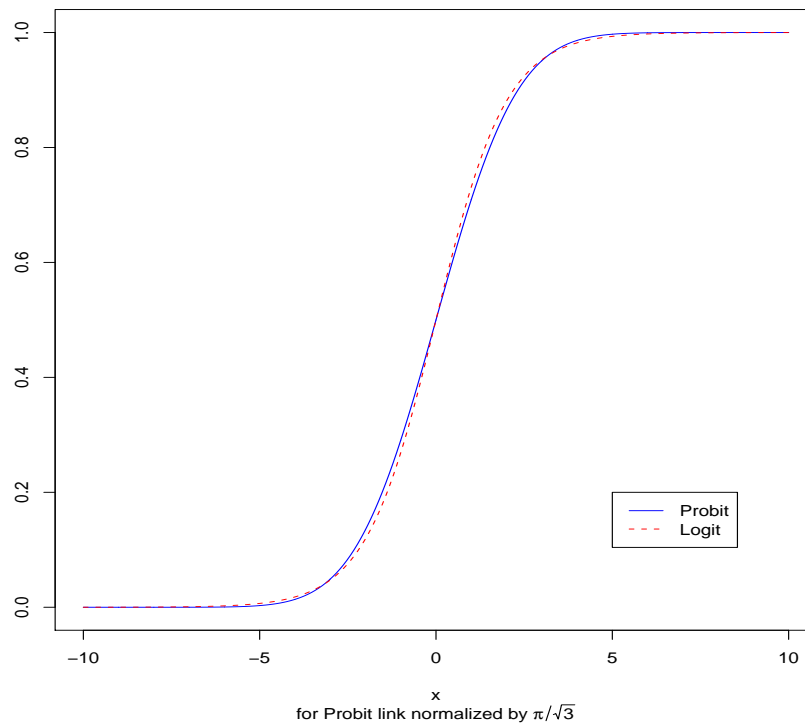


Figure 2: Comparing the Probit link and the Logit link; that is the distribution function of the standard Normal distribution, $\Phi\left(\frac{x}{\pi/\sqrt{3}}\right)$ and distribution function of the Logistic distribution, $\frac{\exp(x)}{1+\exp(x)}$, respectively.

The covariate X has a Binomial distribution with probability of success 0.8 at any time point based on a grid of 100 per time unit. The intercept $\alpha_0(t)$ is $2 - t$.

Denote $X_i^*(t) = X_i(t)/(\pi/\sqrt{3})$. The outcome model (14), using the approximation above,

is based on the following equation for outcome:

$$Y_i(t) = I(\alpha_0^*(t) + \beta_0^* X_i^*(t) + \theta(Z_i(t) - \mu_Z) + \phi_i + \epsilon_i(t) > 0), \quad (15)$$

where $\mu_Z = E(Z_i(t)|X_i(t))$ and where we include the random effect term ϕ_i to model exchangeable correlation on outcome within the same individual as well as the error term $\epsilon_i(t)$. To avoid confounding we incorporate additional covariate $Z_i(t)$ in a perpendicular way to the covariate $X_i(t)$. All three components are mutually independent and Normally distributed. The random effect and the error have mean 0 and variance σ_ϕ^2 and σ^2 , respectively. The covariate $Z_i(t)$ has mean μ_Z and variance σ_Z^2 conditional on $X_i(t)$. A proper marginalization over the random effect, the covariate $Z_i(t)$ and the error term yields

$$\begin{aligned} \beta_0^* &= \beta_0 \sqrt{\sigma^2 + \sigma_\phi^2 + \theta^2 \sigma_Z^2} \\ \alpha_0^*(t) &= \alpha_0(t) \frac{\sqrt{\sigma^2 + \sigma_\phi^2 + \theta^2 \sigma_Z^2}}{\pi/\sqrt{3}} \end{aligned}$$

because based on expression (15)

$$\begin{aligned} E(Y_i(t)|X_i(t)) &= P(\alpha_0^*(t) + \beta_0^* X_i^*(t) + \theta(Z_i(t) - \mu_Z) + \phi_i + \epsilon_i(t) > 0) \\ &= P\left(-\frac{\theta(Z_i(t) - \mu_Z) + \phi_i + \epsilon_i(t)}{\sqrt{\sigma^2 + \sigma_\phi^2 + \theta^2 \sigma_Z^2}} < \frac{\alpha_0^*(t) + \beta_0^* X_i^*(t)}{\sqrt{\sigma^2 + \sigma_\phi^2 + \theta^2 \sigma_Z^2}}\right) \\ &= \Phi\left(\frac{\alpha_0^*(t) + \beta_0^* X_i^*(t)}{\sqrt{\sigma^2 + \sigma_\phi^2 + \theta^2 \sigma_Z^2}}\right). \end{aligned}$$

The Z covariate is a mixture of two Normal distributions, to be specific

$$Z|X \sim (1 - X)N(\mu_{Z1}, \sigma_{Z1}^2) + XN(\mu_{Z2}, \sigma_{Z2}^2).$$

We set $\sigma = 1$, $\sigma_\phi = 0.5$, $\beta_0 = 1$, $\theta = 0.5$, $\mu_{Z1} = 2$, $\mu_{Z2} = 0$, $\sigma_{Z1} = 1$, $\sigma_{Z2} = 2$.

For the sampling-times model, the observation times follow a random-effect Poisson counting process with intensity $\lambda_i(t) = \eta_i \exp\{\gamma_{01} X_i(t) + \gamma_{02} Z_i(t)\}$. Random effect η_i is Gamma distributed with mean $\mu_\eta = 1$ and variance $\sigma_\eta^2 = 0.01$. Thus for each individual the times of observations are positively correlated unless $\sigma_\eta = 0$. Parameter γ_1 we set to 0.2, γ_2 to 0.3.

The censoring variable C is distributed uniformly on the interval $(\tau/2, \tau)$. Parameter τ has the meaning of the median number of observations per individual. Setting τ to 4 and 8

should demonstrate cases of a few and many observations per person. The resulting quantiles of number of observations taken over all individuals and simulations are in Table 5. The parameter assigning weight to time is set to 1 over the entire time span.

Table 5: Quantiles of number of observations per person.

| | minimum | 25% | median | 75% | maximum |
|------------|---------|-----|--------|-----|---------|
| $\tau = 4$ | 1 | 3 | 4 | 5 | 17 |
| $\tau = 8$ | 1 | 6 | 8 | 10 | 25 |

Table 6: Summary statistics of the estimator of β_0 for the simulation studies. We present bias (Bias), sampling standard errors (SSE), sampling mean of estimated standard errors (SEE), coverage probability of the true β_0 (CP) and two measures (M_1, M_2) comparing squared errors between the proposed estimator and the GEE estimator.

| | | Bias | SSE | SEE | CP | M_1 | M_2 |
|-----------|------------|--------|-------|-------|------|-------|-------|
| $n = 20$ | $\tau = 4$ | -0.049 | 0.388 | 0.399 | 0.95 | 2.59 | 4.97 |
| | $\tau = 8$ | -0.078 | 0.258 | 0.289 | 0.96 | 4.31 | 7.47 |
| $n = 50$ | $\tau = 4$ | -0.041 | 0.225 | 0.244 | 0.96 | 5.43 | 10.83 |
| | $\tau = 8$ | 0.039 | 0.164 | 0.168 | 0.95 | 7.69 | 14.35 |
| $n = 200$ | $\tau = 4$ | -0.029 | 0.106 | 0.108 | 0.94 | 11.82 | 21.31 |
| | $\tau = 8$ | -0.023 | 0.079 | 0.081 | 0.94 | 14.20 | 21.16 |
| $n = 500$ | $\tau = 4$ | 0.018 | 0.068 | 0.071 | 0.94 | 16.71 | 22.37 |
| | $\tau = 8$ | -0.009 | 0.051 | 0.053 | 0.95 | 19.11 | 22.79 |

In Table 6 we present bias, sampling standard error of $\hat{\beta}$, SSE, and sampling mean of estimated standard errors, SEE, taken over 1000 simulations. The estimated standard errors are based on a sandwich-form estimator of variance. We also present two measures of squared errors comparison between the two estimation approaches, the proposed biased sampling adjusted estimator and the traditional GEE estimator with independent working assumption. Measure M_1 based on mean of the ratio of empirical mean squared error of estimate (11) of β_0 over empirical mean squared error of GEE estimate of β_0 . Measure M_2 is based on empirical median

of ratios of squared errors, a more robust efficiency estimate motivated by Pitman closeness. We report 95% sampling coverage probability. Because we run 1000 simulations, the precision of the coverage probability is 1.4%. Number of individuals in a sample is set to 20, 50, 200 and 500.

The bias of the IIRR estimates is negligible relative to the sampling standard error. The sampling mean of estimated standard errors is very close to the sampling standard errors, suggesting that the sandwich variance estimator is a good statistic to estimate the variance of the parameter β_0 . We note that in our simulations it always slightly overestimates the true standard errors. The coverage probabilities of the IIRR estimates are also very close to the ideal 95%. The traditional GEE estimator has failed under the biased settings, systematically overestimating the parameter β_0 . We see the inconsistency from the statistics M_1 and M_2 comparing the squared errors between the proposed IIRR estimates and the traditional GEE estimates.

6. Discussion

In survival analysis settings, knowledge of the covariate process over time up to dropout is a standard assumption. In our notation it means knowledge of covariates Z over time if those are time-varying. On the contrary, in longitudinal analysis settings, the usual assumption is that covariates X are known at sampling times only. Our IIRR estimator, combining survival analysis and longitudinal analysis approaches together, requires knowledge of both covariate processes X and Z over time up to dropout, actually only at the data set's sampling times \mathbb{T} . The covariate process X is needed over time because it is used in the stabilized weight $h_0(\cdot)$. Using approximations of the covariate process X or Z cause the IIRR estimators to be biased. However, subcohort sampling techniques such as case-cohort design, pioneered by Prentice (1986), seem promising in keeping consistency of the estimator.

Separation of the sampling-times model and the outcome model as two distinct models enables to perform model checking separately. A range of model checking techniques of the sampling-times model (2) was suggested in Section 4 of Lin et al. (2000) using cumulative sums of residuals based on the process $\{\mathcal{M}(t), t \in [0, \tau]\}$ defined in equation (9). We keep in mind that those residuals are not martingales and thus present a technical challenge to construct formal

tests. Those model checking techniques include both graphical and numerical inspections for functional form of covariates, exponential link function and proportional rates assumptions. For the last we can plot Schoenfeld residuals against time with a fitted smoother, just as described in Grambsch & Therneau (1994) for checking proportional hazard assumption. Also an omnibus test for checking the overall fit of the model was constructed. Lin et al. (2002) in Section 3 provide a guideline on constructing cumulative and moving sums of residuals for longitudinal marginal models. Those should graphically and also numerically, using the Kolmogorov–type supremum test, help to determine the nature of model misspecification. We would like to address similar matter for the outcome model in our future work.

Similarly to H. Lin et al. (2004), we can expand the class of the outcome models from generalized linear models to a broader class, defined as

$$E[Y_i(t)|X_i(t)] = \mu(t, X_i(t); \beta_0) \quad \forall t \in [0, \tau], \quad (16)$$

where the mean of outcome at time t is function of time, covariates at that time and the unknown parameter vector β_0 .

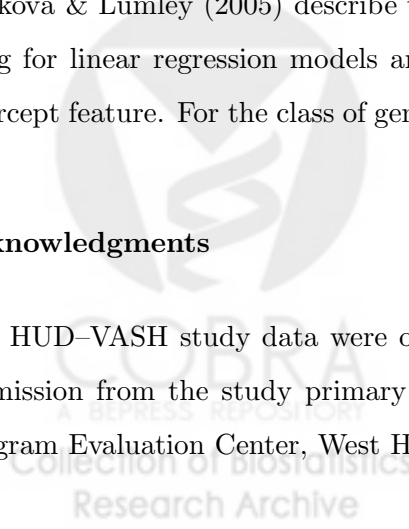
D. Lin & Ying (2001) introduced the integration of counting processes with linear regression models where they allow for an unspecified intercept function $\alpha_0(t)$. Their semiparametric outcome model is

$$E[Y_i(t)|X_i(t)] = \alpha_0(t) + \beta_0^T X_i(t) \quad \forall t \in [0, \tau]. \quad (17)$$

Bůžková & Lumley (2005) describe the extension of Lin and Ying's approach into biased sampling for linear regression models and models with log-links, accommodating the unspecified intercept feature. For the class of generalized regression models this feature seems unattainable.

Acknowledgments

The HUD–VASH study data were obtained from H. Lin at Yale University, New Haven with permission from the study primary investigator R. Rosenheck at Veterans Affairs Northeast Program Evaluation Center, West Haven.



Appendix A: Assumptions

We assume that $(Y_i(\cdot), X_i(\cdot), Z_i(\cdot), N_i^*(\cdot), \xi_i(\cdot))$ are i.i.d. quintuples of random processes over time $t \in [0, \tau]$ for individuals 1 through n . The counting uncensored process of events at the end of follow-up τ , $N_i(\tau)$, is required to be bounded by a constant. Both mean outcome model covariates $X_i(\cdot)$ and sampling times model covariates $Z_i(\cdot)$ need to have bounded total variations by a constant for all individuals $i = 1, \dots, n$. That is $|Z_{ji}(0)| + \int_0^\tau |dZ_{ji}(t)| \leq K$, $j = 1, \dots, g$ and $|X_{ji}(0)| + \int_0^\tau |dX_{ji}(t)| \leq K$, $j = 1, \dots, p$. Total number of observations per individual i , denoted by K_i , is bounded. The weight function $W(\cdot)$ is a difference of two monotone functions, each of which converges to a deterministic function. We denote the limit of $W(\cdot)$ by $w(\cdot)$. We assume that the function $h(\cdot)$ has bounded variation. That is for some $K < \infty$

$$|h(0)| + \int_0^\infty |dh(x)| \leq K.$$

Appendix B: Implementation of the estimation procedure

The estimating procedures developed in this paper can be implemented in S-plus/R with extreme ease. The sampling-times model (2) can be fitted by function *coxph* in package *survival* in order to obtain the estimate of γ_0 . To fit the outcome model (1) we can use the function *glm* from the *stats* package that produces estimates of β_0 by the iteratively reweighted least squares procedure as described in Chambers & Hastie (1991). We only need to specify the inverse weights correctly. The standard errors of the estimate of β_0 can be obtained by bootstrapping or implementing the sandwich estimates provided in the article.

References

- ALBERT, P. S., FOLLMANN, D. A., WANG, S. A., & SUH, E. B. (2002). A latent autoregressive model for longitudinal binary data subject to informative missingness. *Biometrics* **58**, 631–642.
- BŮŽKOVÁ, P. & LUMLEY, T. (2005). Marginal regression modeling under irregular, biased sampling. *Journal of the American Statistical Association* (submitted).
- CHAMBERS, J. M. & HASTIE, T. J. (1991). *Statistical Models in S*. Chapman and Hall Computer Science Series.

- DIGGLE, P. & KENWARD, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics* **43**, 49–93.
- GRAMBSCH, P. M. & THERNEAU, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, 515–526.
- HERNÁN, M. A., BRUMBACK, B. A., & ROBINS, J. M. (2002). Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Statistics in Medicine* **21**, 1689–1709.
- LIANG, K. Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
- LIN, D. Y., WEI, L. J., YANG, I., & YING, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, B Series* **62**, 711–730.
- LIN, D. Y., WEI, L. J., & YING, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics* **58**, 1–12.
- LIN, D. Y. & YING, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* **96**, 103–126.
- LIN, H., SCHARFSTEIN, D. O., & ROSENHECK, R. A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society, Series B* **66**, 791–813.
- LIN, X. & CARROLL, R. J. (2001). Semiparametric regression for clustered data. *Biometrika* **88**, 1179–1185.
- LIPSITZ, S. R., FITZMAURICE, G. M., IBRAHIM, J. G., GELBER, R., & LIPSHULTZ, S. (2002). Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics* **58**, 621–630.
- LITTLE, R. J. A. & RUBIN, D. B. (2002). *Statistical analysis with missing data*. Wiley Series in Probability and Statistics.
- MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized linear models*. Chapman and Hall.
- PEPE, M. S. & ANDERSON, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics - Simulation* **23**, 939–951.

- PREISSER, J. S., GALECKI, A. T., LOHMAN, K. K., & WAGENKNECHT, L. E. (2000). Analysis of smoking trends with incomplete longitudinal binary response. *Journal of the American Statistical Association* **95**, 1021–1031.
- PRENTICE, R. L. (1986). A case cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1–11.
- R DEVELOPMENT CORE TEAM (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.
- ROSENHECK, R., KASPROW, W., FRISMAN, L., & LIU-MARES, W. (2003). Cost-effectiveness of supported housing for homeless persons with mental illness. *Archives of General Psychiatry* **60**, 940–951.
- RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
- SCHARFSTEIN, D. O., ROTNITZKY, A., & ROBINS, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association* **94**, 1096–1120.
- TROXEL, A. B., LIPSITZ, S. R., & HARRINGTON, D. P. (1998). Marginal models for the analysis of longitudinal measurements with nonignorable non-monotone missing data. *Biometrika* **85**, 661–672.
- WU, M. C. & CARROLL, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188.

