Published in final edited form as:

Biometrics. 2010 September ; 66(3): 726–732. doi:10.1111/j.1541-0420.2009.01326.x.

A Score Test for Association of a Longitudinal Marker and an Event with Missing Data

Dianne M. Finkelstein^{1,*}, Rui Wang¹, Linda H. Ficociello², and David A. Schoenfeld¹

¹Massachusetts General Hospital and Harvard University, Biostatistics Unit, 50 Staniford Street, Boston MA 02114

²Joslin Diabetes Center, 1 Joslin Place, Boston MA 02215

Summary

Often clinical studies periodically record information on disease progression as well as results from laboratory studies that are believed to reflect the progressing stages of the disease. A primary aim of such a study is to determine the relationship between the lab measurements and disease progression. If there were no missing or censored data, these analyses would be straightforward. However, often patients miss visits, and return after their disease has progressed. In this case, not only is their progression time interval-censored, but their lab test series is also incomplete. In this paper, we propose a simple test for the association between a longitudinal marker and an event time from incomplete data. We derive the test using a very intuitive technique of calculating the expected complete data score conditional on the observed incomplete data (CEST). The problem was motivated by data from an observational study of patients with diabetes.

Keywords

EM algorithm; Interval censored failure time data; PRO logistic model; Random effects model; Conditional Expected Score Test (CEST)

1. Introduction

In studies of chronic diseases such as cancer and AIDS, patients are monitored for clinical or laboratory measurements (or markers) as well as events that are known to be associated with declining health and an increased risk of death. Understanding the pattern of these markers and events is important for the clinical management of individual patients, for deepening understanding of the natural history of the disease, and for the design and analysis of clinical trials for new therapies. The statistical analysis of such data is common, with standard methods such as the Cox proportional hazards model with a time-varying covariate. However, the analysis may be complicated by the fact that patients miss visits, resulting in incomplete information either on when an event occurred (*interval censored data*) and/or on the value of serially collected clinical or laboratory measurements that are important for prediction of the event. The simple approach of analyzing only complete data can be biased and thus it is necessary to appropriately handle the missing data.

*dfinkelstein@partners.org.

Supplemental Materials The Web Appendix with calculation of variance of the test statistic referenced in Section 2.5, and the tables with results of simulations for size and power of the test referenced in Section 3 are available under the Paper Information link at the Biometrics website <http://www.biometrics.tibs.org>.

A recent example of this type of study involved a group of patients at the Joslin Diabetes Center who were followed for evidence of change in albumin excretion, which is an indicator of diabetic kidney disease (proteinuria) (Perkins, et al. 2003, Ficociello, et al. 2007). The investigators were interested in determining laboratory markers that could indicate which patients were at elevated risk for renal (kidney) disease. At scheduled visits, several lab tests were performed including a screening for presence of an excess of serum proteins in the urine, or proteinuria, which can be a sign of renal disease. One aim of the study was to determine whether blood sugar control was associated with an elevated risk of proteinuria. As a measure of diabetes control, levels of glycohemoglobin A1c (A1c) were recorded from the blood specimen. The investigators were interested in whether patients who had poorer diabetes control were at greater risk for subsequent proteinuria (renal disease). This is depicted in the top row of Figure 1. There were many patients who missed one or more clinic visits, and returned with evidence of proteinuria. When patients missed clinic visits, then neither A1c nor urine protein levels (used to determine the occurrence of the event of proteinuria) were available. Thus, as indicated in the bottom row of Figure 1, if a patient missed a visit at time j , and returned with proteinuria at time $j + 1$, then the A1c for visit j was unavailable for predicting the onset of proteinuria at time $j + 1$; also the onset of proteinuria would be interval censored. Therefore, a routine Cox model with a time-varying covariate could not be used for the analysis.

These data arise in AIDS studies as well. For example, the AIDS Clinical Trials Group (ACTG) study 181 (Finkelstein et al., 2002) periodically monitored the blood and urine of patients for evidence of CMV disease. The investigators were interested in whether the CD4 count from the blood sample could predict risk of future CMV disease. A missing lab visit resulted in both missing CD4 and missing CMV marker data which provide evidence of CMV disease.

Methodology has been proposed for the estimation and regression analysis of the effect of baseline predictors on interval-censored failure time data (see Finkelstein, 1986 and Goggins et al. 1998). In addition, there are methods for regression of failure time data with missing longitudinal covariates, such as Goggins et al. (1999), which considered the case of a missing time-varying binary covariate (resulting in interval censored covariate data). Others have developed methods for analysis of data with missing covariates for the case when the failure times are right-censored, relying on joint modeling of the covariate and failure time distributions; these are well summarized in Hogan and Laird (1997). Computational methods for these models can be intensive, particularly when a Bayesian approach is taken (Guo and Carlin, 1994). While these methods are appropriate for right censored failure time data, they cannot be readily extended to interval censored data. The goal of this paper is to develop a simple and intuitive test for the relationship between the longitudinal covariate and a failure time when the time-varying covariates can be missing and the failure times are interval censored.

To develop this test we apply a powerful yet simple technique for deriving a score test in the presence of incomplete data which utilizes the fact that the score based on incomplete data is equal to the expected value of the complete data score conditional on the observed data. This allows a particularly simple derivation of the test statistic. This fact was noted by Efron in the discussion of Dempster, Laird and Rubin (1977). We call this the conditional expected score test (CEST) and note that while the longitudinal/event time application is an illustration of this technique, it provides a paradigm for developing tests for a broader class of missing data problems.

In Section 2, we will first introduce a score test for the relationship between an observed longitudinal covariate and an event for the case when all data are completely observed. Next

we will introduce the random effects model to handle the problem that the covariate value may be missing for some subjects at some times. We then apply CEST to extend the score test to handle the case where some visits are missed, resulting in interval censored event (and missing covariate) data. We later extend the test again to the case where the covariate is measured with error, and the trajectory of the covariate is the appropriate predictor of failure rather than the observed covariate. In Sections 3 and 4, we provide simulations and an application of our methods to analyze a study of renal disease in type 1 diabetes. We discuss possible extensions, etc. in Section 5.

2. Methods

2.1 Notation and Model

Suppose that we completely observe the data for all subjects. We will let T_i be the time patient i has an event, for $i = 1, \dots, n$, and let t_j be pre-determined equally spaced clinic visits at which the patients are monitored, for $j = 1, \dots, M$. Let δ_{ij} be an indicator of whether or not patient i had failed since the previous clinic visit at t_{j-1} , Y_{ij} be an indicator of whether or not they were in follow-up (and at risk) at time j , and let Z_{ij} be their covariate at the previous clinic visit which will be used to predict the failure in the j^{th} interval.

To model the relationship between the covariate at the previous visit and subsequent failure, we will utilize a generalized person-year approach described in Cupples et al (1988), by which observations over multiple time intervals are pooled into a single sample, and logistic regression is used to relate the risk factors to the occurrence of the event. When grouping intervals between exams are short, this Pooling Repeated Observations method (PRO) is asymptotically equivalent to grouped Cox model for a time-dependent covariate (D'Agostino, 1990). In our application, the response variable of the logistic model will indicate failure as a function of the covariate at the previous time, z_{ij} :

$$\text{logit } p_{\beta}(z_{ij}) = \log \frac{p_{\beta}(z_{ij})}{1 - p_{\beta}(z_{ij})} = \alpha_j + \beta' z_{ij} \quad (1)$$

where $p_{\beta}(z_{ij}) = p(\delta_{ij} = 1 | z_{ij}, T_i > t_{j-1})$, which is the conditional probability of observing an event in the j^{th} interval given the individual was event-free through time t_{j-1} and z_{ij} is the (possibly vector valued) value of the covariate measured at time t_{j-1} . For the PRO model, the intercept $\alpha_j = \alpha$ is assumed to be constant across time, as suggested by the investigator. This can be relaxed as we will describe in the discussion of this paper. In all our applications, we will assume that the failure status is “absorbing” and note that patients are no longer under observation after they fail or complete all the visits. Also, we will assume that determination of δ_{ij} is made prior to the measurement of Z at time j .

We can now write the log-likelihood for the data as:

$$\sum_{i=1}^n \sum_{j=1}^M \left[-Y_{ij} \log(1 + e^{\alpha + \beta' z_{ij}}) + \delta_{ij} (\alpha + \beta' z_{ij}) \right] \quad (2)$$

The numerator of the score test for the hypothesis $H_0 : \beta = 0$ (for complete data) would be given by:

$$U = \sum_i U_i = \sum_{i=1}^n \sum_{j=1}^M [\delta_{ij} z_{ij} - Y_{ij} z_{ij} \hat{p}] \quad (3)$$

where

$$\hat{p} = \frac{\sum_i \sum_j \delta_{ij}}{\sum_i \sum_j Y_{ij}} \quad (4)$$

is the MLE for $p = p_0 = \frac{e^\alpha}{1+e^\alpha}$, the probability of failure at each time under $H_0 : \beta = 0$.

2.2 Random Effects Model for the Covariate Data

Now suppose that the value of the covariate is missing for some subjects at some times. We will assume the covariate z_{ij} follows a linear growth curve model with random effects as in Wulfsohn and Tsiatis (1997). More specifically, we assume

$$z_{ij} = \theta_{0i} + \theta_{1i} t_{j-1} + e_{ij},$$

where the error e_{ij} is from a $N(0, \sigma_e^2)$ distribution, and $cov(e_{ij}, e_{ij'}) = 0$ for $j \neq j'$, and that the error is independent of the random intercept and slope, θ_{0i} and θ_{1i} . In addition, we assume

that $\theta_i \sim MN(\theta, V)$, where $\theta_i = (\theta_{0i} \ \theta_{1i})'$, $\theta = (\theta_0 \ \theta_1)'$, and $V = \begin{pmatrix} \sigma_{00} & \sigma_{01} \\ \sigma_{01} & \sigma_{11} \end{pmatrix}$

We will assume that the measured covariate value predicts the event, but it may be missing for some people at some observation times, and call this Model O (Observed). Later in the paper, we will consider an alternate paradigm for which we will assume that the covariate is measured with error, and that the trajectory (from the random effects model) predicts the failure, called Model T (trajectory). In this case, the true value of the covariate is unobserved for all subjects at all times, and these values are treated as missing data. In both cases, as we will show later, we can use the predicted values obtained from the above random effects model to impute the missing covariates.

2.3 Derivation of Score with Missing Data Using CEST

We now consider the problem that the event data can be censored into intervals bracketed by the missing visits. Suppose that the i th subject missed visits after time t_{Li} , came back at t_{Ri} , and that $\delta_{iLi} = 0$ and $\delta_{iRi} = 1$. In this case, δ_{ij} is missing for $L_i + 1 \leq j \leq R_i - 1$. Also because the event is an absorbing state, if any subject misses a visit but returns at t_k still event-free, then we assume $\delta_{ij} = 0$ for all $j \leq k$. Let M_i be the index of the last time subject i was

observed. We define $\delta_{i\cdot} = \sum_{j=1}^{M_i} \delta_{ij}$ which is an indicator of whether the subject ever failed, or left follow-up without an event. If the subject failed while in follow-up, then $\delta_{i\cdot} = 1$ and $R_i = M_i$. If the subject has not failed by their last visit at $M_i = L_i$, then they are right censored, $\delta_{i\cdot} = 0$, and $R_i = \infty$. We note that whenever a subject misses a visit, the covariate measured at that time (used to predict the subsequent event) will be missing as well. However, because we can assume that the covariate follows the linear growth curve model discussed in the previous section, it will be possible to use the aforementioned approach (described in

Section 2.2) in a straightforward way to handle missing observed covariate values. We will assume that our data for each subject consists of values of z up until M_i , their last observed time (although some may be missing). Also, the failure time information is captured by time indices L_i, R_i ; if a subject misses a visit and returns without the event, then their event status is assumed unchanged during the interval of the visit missed. We assume that any missing data is missing at random (MAR).

We will derive the score test for the association between the covariate and event time first directly from the likelihood. We will then show the result could be derived by utilizing the technique we have called the conditional expected score test (CEST).

Likelihood with missing data—We note that the log of the likelihood for the partially missing data can be written:

$$\begin{aligned} \ell(\alpha, \beta, \theta, V, \sigma_e^2) \\ = \sum_{i=1}^n \log \left[\int_{\mathcal{M}_i} \sum_{k \in (L_i, M_i]} f(z_{i1}, \dots | \theta_i, \sigma_e^2) h(\theta_i | \theta, V) \right. \\ \left. \times g(t_k, \delta_i | z_{i1}, \dots, \theta_i, \alpha, \beta) d\mathcal{M}_i \right] \end{aligned} \quad (5)$$

where $M_i = (\theta_{i0}, \theta_{i1}, z_{ij1}, \dots)$ are missing for subject i and

$$\begin{aligned} f(z_{i1}, \dots | \theta_i, \sigma_e^2) &= \prod_{j=1}^{L_i} (2\pi\sigma_e^2)^{-1/2} \exp \left\{ -\frac{(z_{ij} - \theta_{0i} - \theta_{1i}t_j)^2}{2\sigma_e^2} \right\} \\ h(\theta_i | \theta, V) &= (2\pi |V|)^{-1/2} \exp \left\{ -(\theta_i - \theta)' V^{-1} (\theta_i - \theta) / 2 \right\} \\ g(t_k, \delta_i | z_{i1}, \dots, \theta_i, \alpha, \beta) &= [\exp(\alpha + \beta z_{ik})]^{\delta_i} \prod_{j=1}^k \left[\frac{1}{1 + \exp(\alpha + \beta z_{ij})} \right] \end{aligned}$$

where k is an index of a time interval in $(L_i, M_i]$. Note that if $\delta_i = 0$, then g is only evaluated at $k = M_i$, in which case it is the probability of no failure seen in a patient up to time M_i .

The score for subject i can be written as A_i/B_i , where

$$\begin{aligned} A_i &= \int_{\mathcal{M}_i} \sum_{k \in (L_i, M_i]} f(z_{i1}, \dots | \theta_i, \sigma_e^2) h(\theta_i | \theta, V) \\ &\quad \times g(t_k, \delta_i | z_{i1}, \dots, \theta_i, \alpha, \beta) \\ &\quad \times \frac{\partial}{\partial \beta} \log g(t_k, \delta_i | z_{i1}, \dots, \theta_i, \alpha, \beta) d\mathcal{M}_i \end{aligned} \quad (6)$$

$$\begin{aligned} B_i &= \int_{\mathcal{M}_i} \sum_{m \in (L_i, M_i]} f(z_{i1}, \dots | \theta_i, \sigma_e^2) h(\theta_i | \theta, V) \\ &\quad \times g(t_m, \delta_i | z_{i1}, \dots, \theta_i, \alpha, \beta) d\mathcal{M}_i \end{aligned} \quad (7)$$

Under $H_0 : \beta = 0$,

$$\begin{aligned} &\frac{\partial}{\partial \beta} \log g(t_k, \delta_i | z_{i1}, \dots, \theta_i, \alpha, \beta) \\ &= \sum_{j=1}^k \left[\delta_i \cdot I(j=k) - I(j \leq k) \cdot \frac{e^\alpha}{1+e^\alpha} \right] z_{ij} \end{aligned}$$

Note that when $\beta = 0$, g does not depend on z or θ so both (6) and (7) can be factored. Let $g_k = g(t_k, \delta_i, \alpha)$ we can write the individual score contribution A_i/B_i as:

$$\begin{aligned}
 A_i/B_i &= \sum_{k \in (L_i, M_i]} \sum_{j=1}^k \left[\frac{\{\delta_i \cdot I(j=k) - p \cdot I(j \leq k)\} g_k}{\sum_{m \in (L_i, M_i]} g_m} \right. \\
 &\quad \times \left. \frac{\int_{\mathcal{M}_i} z_{ij} f(z_{i1}, \dots, \theta_i, \sigma_e^2) h(\theta_i | \theta, V) d\mathcal{M}_i}{\int_{\mathcal{M}_i} f(z_{i1}, \dots, \theta_i, \sigma_e^2) h(\theta_i | \theta, V) d\mathcal{M}_i} \right] \\
 &= \sum_{k \in (L_i, M_i]} \sum_{j=1}^k \left[\frac{\{\delta_i \cdot I(j=k) - p \cdot I(j \leq k)\} g_k}{\sum_{m \in (L_i, M_i]} g_m} E(z_{ij}) \right] \\
 &= \sum_{j=1}^{M_i} \sum_{k=L_i+1}^{M_i} \left[\frac{\{\delta_i \cdot I(j=k) - p \cdot I(j \leq k)\} g_k}{\sum_{m \in (L_i, M_i]} g_m} E(z_{ij}) \right] \\
 &= \sum_{j=1}^{M_i} E[\delta_{ij} - p Y_{ij} | T_i \in (t_{L_i}, t_{R_i}]] E_i(z_{ij})
 \end{aligned} \tag{8}$$

where $E(z_{ij})$ is the expectation of z_{ij} conditional on all observed z_i under the random effects model we assumed in the previous section (2.2.) and $p = e^\alpha (1 + e^\alpha)^{-1}$. To calculate the test statistic in (8), we need to calculate the expected values for the missing data which we do in the next section. We also need to obtain the MLEs for the nuisance parameters $(\alpha, \sigma_e^2, \theta, V)$ under $H_0 : \beta = 0$. It can be shown that under H_0 , (5) can be written as $\ell(\alpha) + \ell(\theta, V, \sigma_e^2)$, where $\ell(\alpha)$ only involves failure status and $\ell(\theta, V, \sigma_e^2)$ only involves covariates. Thus, the estimation of α only uses data from failure status and the estimation of (σ_e^2, θ, V) only depends on observed covariate data.

Conditional Expected Score Test (CEST)—We now note that the score presented in (8) could actually have been derived from (3) using a general principle for derivation of the score test from missing data that was noted by Efron in the discussion of Dempster, Laird and Rubin (1977), namely that the score based on incomplete data is the expected value of the complete data score conditional on the observed data. More formally (using notation of Dempster et al, 1977), if $\partial \mathcal{L}^x(\phi)$ is the Fisher score function based on the complete data set x , and $\partial \mathcal{L}^y(\phi)$ is the score function based on a statistic $y(x)$ (which in our case is the incomplete data y), then

$$\partial \mathcal{L}^y(\phi) = E_\phi(\partial \mathcal{L}^x(\phi) | y) \tag{9}$$

We call this method the conditional expected score test (CEST) and note that use of this for the case of missing longitudinal and failure time data simplifies an otherwise complex derivation. We only need to obtain the expectations for the data that are missing. This approach can be applied to a broad class of tests in the context of missing data.

Expectation of missing data—To calculate the conditional expectations in (8), we note first that under $H_0 : \beta = 0$, (δ_{ij}, Y_{ij}) is independent of $(\mathbf{z}_i, \theta_{0i}, \theta_{1i})$. We need to obtain the expectation of δ and Y conditional on the observed data, $E(\delta_{ij} | \delta, \mathbf{Y}, \beta = 0)$ and $E(Y_{ij} | \mathbf{Y}, \beta = 0)$ for each subject i who failed during follow-up and whose event is censored in $(L_i, R_i]$. To do this, we obtain the values of the probability of the event occurring at each of the visits in the censoring interval. These quantities are calculated under the null hypothesis $H_0 : \beta = 0$.

$$\begin{aligned}
E(\delta_{ij}|\delta, Y, \beta=0) &= w_{dij} \\
&= p(T_i = t_j | t_{L_i} < T_i \leq t_{R_i}, \delta, Y, \beta=0) \\
&= \frac{(1-p)^{j-1}}{\sum_{k=L_i+1}^{R_i} (1-p)^{k-1}}
\end{aligned} \tag{10}$$

for j such that $t_j \in (L_i, R_i]$ and 0 otherwise and

$$\begin{aligned}
E(Y_{ij}|\delta, Y, \beta=0) &= w_{rij} \\
&= p(T_i \geq j | t_{L_i} < T_i \leq t_{R_i}, \delta, Y, \beta=0) \\
&= 1 - \sum_{l=1}^{j-1} w_{dil}
\end{aligned} \tag{11}$$

Note that $w_{rij} = 1$ for $j \leq L_i$ and is zero for $j > R_i$. Also note that for a subject who did not fail while under observation (from times 1, 2, ..., M_i), $w_{dij} = 0$ and $w_{rij} = 1$ for $j = 1, \dots, M_i$. To

obtain the MLE for $p = \frac{e^\alpha}{(1+e^\alpha)}$ under H_0 , we solve the observed data score equation for α :

$$\sum_{i=1}^n \sum_{j=1}^{M_i} (w_{dij} - w_{rij}p) \tag{12}$$

This estimate can be obtained by substituting the expected values of Y and δ when these are unknown for interval censored subjects in (4); thus $\hat{p}$, is the value of p that satisfies the self consistency equation (Efron 1967):

$$p = \frac{\sum_i \sum_j w_{dij}}{\sum_i \sum_j w_{rij}} \tag{13}$$

We do this by an EM type algorithm. We start with the value of p calculated from complete data only, and iterate until convergence.

To calculate $E(z_{ij})$, we note that:

$$E(z_{ij}) = \begin{cases} z_{ij} & \text{if } z_{ij} \text{ is observed} \\ \hat{\theta}_{0i} + \hat{\theta}_{1i} t_{j-1} & \text{otherwise} \end{cases} \tag{14}$$

The observed data score can be calculated by substituting (10) and (11) into (8):

$$U = \sum_{i=1}^n \sum_{j=1}^{M_i} (w_{dij} - \hat{p} \cdot w_{rij}) \cdot \bar{z}_{ij} \tag{15}$$

where $\tilde{z}_{ij} = E(z_{ij})$, the expectation calculated as noted above. The exact and right censored individuals contribute the same terms to (15) as they did in (3), while the contribution from each individual i interval censored between $(L_i, R_i]$ is:

$$U_i = \sum_{j=1}^{M_i} (w_{dij} - \hat{p} \cdot w_{rij}) \tilde{z}_{ij} = \sum_j \tilde{z}_{ij} - \hat{p} \sum_{k < j} \tilde{z}_{ij} w_{dij}. \quad (16)$$

Note that the score in (16) can be viewed as a sum of “observed-expected”. In this case, each interval-censored subject contributes to the “observed” value (the first term in (16)) a weighted average of their estimated true covariate values from times j at which their failure could have occurred, with a weight equal to the probability the failure was at that time. They contribute to the “expected” value (the second term in (16)) a weighted average of their estimated true covariate values at the times k they were under observation with a weight equal to the probability that the failure was at or after the time k . The numerator of the score test $U = \sum_i U_i$ can also be written

$$U = \sum_j \left[\sum_{i \in D_j} w_{dij} \tilde{z}_{ij} - \hat{p} \cdot \sum_{i \in R_j} w_{rij} \tilde{z}_{ij} \right]. \quad (17)$$

where D_j is the set of individuals who had (or may have had) an event at time j , and R_j is the set of individuals who are or *may have been* under observation just before time j , and for exact failures, $w_{dij} = \delta_{ij}$ and $w_{rij} = Y_{ij}$. Note that for an individual censored into $(L_i, R_i]$, they will contribute terms to the score at each time through R_i . Under the independence assumption, the scores, conditional on z are independent and the U has mean zero.

Thus, to calculate the score in (17), we first estimate $\hat{p}$ from (13) as well as the weights w_{rij} and w_{dij} . We then calculate $\tilde{z}_{ij}$ under the random effects model using all observed for all subjects and substitute these values into (17).

2.4 Covariate measured with error

We note that our method could be easily applied to the case where a covariate was measured with error, and the trajectory (from the random effects model) predicts the failure. In this case, we replace z_{ij} by $\theta_{0i} + \theta_{1i}t_{j-1}$ and g becomes:

$$g(t_k, \delta_i | z_{i1}, \dots, \theta_i, \alpha, \beta) = \exp[\alpha + \beta(\theta_{0i} + \theta_{1i}t_{k-1})] \times \prod_{j=1}^k \left[\frac{1}{1 + \exp[\alpha + \beta(\theta_{0i} + \theta_{1i}t_{j-1})]} \right].$$

The argument in the previous section goes through, and the individual score contribution becomes:

$$\sum_{j=1}^{M_i} E[\delta_{ij} - pY_{ij} | T_i \in (t_{L_i}, t_{R_i}]] E[\theta_{0i} + \theta_{1i}t_{j-1}]$$

When the expectation is taken with respect to θ_i , conditional on all the z , the score becomes the same as in (15):

$$U = \sum_{i=1}^n \sum_{j=1}^{M_i} (w_{dij} - \hat{p} \cdot w_{rij}) \cdot \tilde{z}_{ij}$$

where now $\tilde{z}_{ij} = \hat{\theta}_{0i} + \hat{\theta}_{1i}t_{j-1}$

Thus, to calculate the score in (15) in this case, we estimate $\hat{p}$ from (13) as well as the weights w_{rij} and w_{dij} as before. We then calculate the random effects model using all observed covariates for all subjects. $\tilde{z}_{ij}$ is then calculated from the random effects model and substituted into (15).

2.5 Variance of the score test

The variance of U in (15) could be calculated using the bootstrap method. Alternatively, we can obtain the variance using standard likelihood theory where we first calculate the empirical Fisher information matrix using individual scores and then invert this matrix and

pick the appropriate element. More specifically, let $\Gamma_i = (\Lambda'_i, \pi_i, U_i)'$, where

$\pi_i = \sum_j (\hat{w}_{dij} - \hat{p} \cdot \hat{w}_{rij})$, $U_i = \sum_j (\hat{w}_{dij} - \hat{p} \cdot \hat{w}_{rij}) \cdot \tilde{z}_{ij}$, and Λ_i denote the empirical individual scores corresponding to the parameters in the random effects model (see Appendix in Supplemental Materials). The Fisher information can be approximated by:

$$M = \sum_i \Gamma_i \Gamma_i' - \frac{1}{n} \left(\sum_i \Gamma_i \right) \left(\sum_i \Gamma_i \right)' \quad (18)$$

Then the variance of U_i will be approximately equal to the reciprocal of the (8,8) element of the inverse of the matrix M . The hypothesis test of $H_0: \beta = 0$ can be performed using Rao's efficient score test by comparing $U / \sqrt{V}$ to a normal distribution. We note that the formula given for Model O (where observed covariate is used when available) work also for Model T (covariate measured with error so trajectory is used). The only difference is the value of the covariate that is used. All proofs for the score and its variance work for both models.

3. Simulation Study

In our simulation study, we generated data similar to the diabetes data set used in our diabetes example. For the simulations we used a random effects model; the variances of the intercept and slope were 1.57 and 0.29, respectively, and their correlation was 0.23. The error term had a variance of 2.7. The random effects were generated using the estimates obtained from the diabetes data set and the events were based on the logistic model with intercept zero and varying slopes. We created data with five visit times. We randomly generated missing visits with a probability of 0, .25 and .5. We ran simulations with p varying from a mean value (over the five times) of .2 to .6, and the slope of the random effects model varying from 0 to 2. We considered both the "observed" model where the covariate is filled in with a value predicted by the random effects model only when it is missing (Model O) and the "trajectory" model when the covariate is assumed to be measured with error, so the trajectory from the random effects model is used as the covariate (Model T). We used sample sizes of 200 and 500. Simulations were repeated 2000 times. The test size was satisfactory, ranging from .04 to .07 for $n = 500$. The 95% confidence bounds on this should have been $\pm 1\%$. The size was similar regardless of missingness,

model, or parameters of the random effects model. The test was slightly anti-conservative, due to an underestimate of the variance of the test. This was more pronounced, as expected, when a smaller sample size (200) was used. We used the asymptotic variance for simulations, but it may be more appropriate to use a more robust variance (such as the bootstrap variance). We assessed power of the test at $\beta = .27$. These results are displayed in tables provided in the Appendix in the Supplemental Materials of this paper. In general, the power of the proposed test increases as p , or the slope in the random effects model increases; and decreases with increasing missingness. Results are similar for the two models. We also performed the simulations using complete cases only, and found the test to be biased away from the null. This is due to the fact that the patients who fail earlier are less likely to be missing before they fail, and if there is a change in the covariate with time, it will appear to be significantly related to the event even under the null.

4. Analysis of Renal Disease in Type 1 Diabetes

Elevation in urinary albumin excretion (to between 30 and 299 ug/min) marks the early stages of diabetic nephropathy termed microalbuminuria. This stage is an early marker of increased risk of advanced diabetic nephropathy, indicating a need for clinical intervention to prevent progression to the more serious state of proteinuria (albumin > 300 ug/min). Microalbuminuria can regress to normal in up to half of the patients.

The Joslin Study of the Natural History of Microalbuminuria in Type 1 Diabetes has been a 10 year observational study of a clinic-based population of young adults (Perkins et al., 2003, Ficociello et al., 2007). This study included a cohort of 381 patients who had evidence of microalbuminuria identified during the 2-year baseline observational period. All patients were scheduled for a baseline and at least four biannual visits at which they underwent clinical examination, and provided urine and blood specimens. Albumin excretion rate and HbA1c measurements were recorded. A1c tends to track over time and its trajectory seems to predict failure (see Figure 2). This study was focused on monitoring early changes in kidney health in type 1 diabetes patients. One aim was to determine the relationship between diabetes control and risk of progressive kidney disease. Hemoglobin A1c (HbA1c) is an average of blood glucose over the past 2-3 months, with high levels associated with poor diabetes control. The investigator was interested in whether previous A1c would predict subsequent onset of proteinuria. Although patients were scheduled for the clinic and lab visits every 2 years, many patients had missed or delayed visits, resulting in missing observations for both the HbA1c and the albumin level that would mark progression in kidney disease. As indicated in Figure 1, if a patient missed visit j , and returned with proteinuria at time $j + 1$, then the A1c for visit j was unavailable for predicting the onset of proteinuria, and the onset of proteinuria would be interval censored, and thus even in a person-years approach analysis, this patient would not have data to contribute to the risk analysis for times j or $j + 1$. The data set has 136 subjects who had complete data at all clinic visits and never failed during follow-up, 46 who had complete data up until they failed, 181 who did not fail during follow-up and left the study early, and 18 subjects who were interval censored on their failure (missed visits and returned having failed).

An initial analysis of these data was performed by the medical investigator. He had chosen to use a person-years approach, by applying the pooled logistic regression (PRO) approach (Cupples et al., 1988, D'Agostino et al., 1990) to simplify the analysis of the longitudinal missing data. He said it was reasonable to assume that only measurements at the most recent previous clinic visit would impact the diagnosis of kidney disease progression at the current visit. Missing observations were left out (complete cases only). This resulted in 1353 subject-visits. For this analysis, using the trajectory model, yielded a test statistic of $Z = -7.25$, and when we used the observed value of the covariate, $Z = -6.35$. Next we

performed the analysis using our proposed method with imputed values for A1c (under random effects model) regardless of whether the visit was missed (assuming that A1c is measured with error), resulting in $Z = -7.07$. When we only used the imputed A1c for intervals for which it was missing, $Z = -6.03$. Using the bootstrap variance, the value of the test statistic in this latter case was $Z = -5.10$, confirming the fact that the test statistic based on the empirical variance is anti-conservative. Thus, A1c is highly associated with risk for kidney disease progression, although the association is less than that found on complete observations only. For this application, the test statistic was highly significant in all approaches. Our computer program will be made available on our website <http://hedwig.mgh.harvard.edu/biostatistics>.

5. Discussion

We were interested in testing for the relationship of the time-varying covariate and the failure when some data were missing, resulting in interval censored failure time and time-varying covariates. We note that for our data, there were only five prescribed visit times at which failure could be determined (discrete failure time), but had the event been recorded in continuous time, it would have been possible to apply these methods by allowing the logistic parameter α to take on distinct values at each of the steps of the Turnbull (1976) empirical estimator. If the sample size and number of failures were sufficient, then the test associated with the logistic PRO model would be asymptotically equivalent to the proportional hazards regression model. (D'Agostino et al, 1990).

The development of our test was simplified by the use of the conditional expected score test (CEST). This approach could be used for other applications as well, as we have demonstrated in recent work (Finkelstein et al, 2008). There are several potential extensions of our methods. First, we made the assumption in this analysis that the censoring was independent. It would be useful to extend this approach to handle dependent censoring such as would be the case if a patient died prior to the event of proteinuria (see Robins and Finkelstein, 2000 and Finkelstein, Goggins and Schoenfeld, 2002). For the diabetes data discussed above, the investigators indicated that only the most recent value of A1c was likely to impact the event of interest, and that the distribution of A1c over time was unknown. To extend this to a model in which the history of A1c predicts failure, one could remove the stationarity condition (constant intercept in the logistic model). This would remove a source of the bias described above for using complete cases only. This more general model could also accommodate the situation when the pre-determined clinic visits were not evenly spaced in time. We could also introduce additional covariates in the model for $p(t|z)$ such as one that captures the slope or previous A1c values. We used a random effects model for the covariate function. Alternatively, if the number of visits is small, we could fit a multivariate normal model. In addition, other methods are available for imputing values for the covariate such as obtaining the value from PROC MI in SAS which assumes a multivariate normal model for z .

Finally, we note that we have focused here on developing a score test. If estimation were the focus of the analysis, we could have used a full likelihood approach using NLMIXED as described in Vonesh, Greene and Schluchter (2006).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to thank Drs. Adrienne Cupples and Andrzej Krolewski for assistance with the diabetes data analysis, and Manor Ashkenazi and Peter Lazar for programming assistance. This work was partially funded by grants from the NIH: HD05520, CA 74302, CA075971, and DK-41526.

References

- Cupples LA, D'Agostino RB, Anderson K, Kannel WB. Comparison of baseline and repeated measure covariate techniques in the Framingham Heart Study. *Statistics in Medicine*. 1988; 7:205–218. [PubMed: 3353604]
- D'Agostino R, Lee ML, Belanger A, Cupples A, Anderson K, Kannel W. Relation of pooled logistic regression to time dependent Cox regression analysis: The Framingham Heart Study. *Statistics in Medicine*. 1990; 9:1501–15. [PubMed: 2281238]
- Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the *EM* algorithm. *Journal of the Royal Statistical Society B*. 1977:1–38.
- Efron, B. The two sample problem with censored data. *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.*; Univ. of Calif. Press; 1967. p. 831-853.
- Ficociello LH, Perkins BA, Silva KH, Finkelstein DM, Ignatowska-Switalska H, Gaciong Z, Cupples LA, Aschengrau A, Warram JH, Krolewski AS. Determinants of progression from microalbuminuria to proteinuria in patients who have type 1 diabetes and are treated with angiotensin-converting enzyme inhibitors. *Clin J Am Soc Nephrol*. 2007; 2(3):461–9. [PubMed: 17699452]
- Finkelstein DM, Goggins WB, Schoenfeld DA. Analysis of failure time data with dependent interval censoring. *Biometrics*. 2002; 58:298–304. [PubMed: 12071402]
- Finkelstein DM. A proportional hazards model for interval-censored failure time data. *Biometrics*. 1986; 42:845–54. [PubMed: 3814726]
- Finkelstein, DM.; Rajicic, N.; Wang, R.; Schoenfeld, D. Analysis of longitudinal and response/failure time data from an observational sStudy with missing observations, survival analysis workshop; Columbia MO. October 18, 2008; 2008.
- Goggins WB, Finkelstein DM, Schoenfeld DA, Zaslavsky AM. A Markov chain Monte Carlo EM algorithm for analyzing interval censored data under the Cox proportional hazards model. *Biometrics*. 1998; 54:1498–1507. [PubMed: 9883548]
- Goggins WB, Finkelstein DM, Zaslavsky A. Applying the Cox proportional hazards model when the change time of a binary time-varying covariate is interval-censored. *Biometrics*. 1999; 55:445–51. [PubMed: 11318198]
- Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*. 2004; 58:16–24.
- Hogan JW, Laird NM. Model-based approaches to analyzing incomplete longitudinal and failure-time data. *Statistics in Medicine*. 1997; 16:259–272. [PubMed: 9004396]
- Perkins BA, Ficociello LH, Silva KH, Finkelstein DM, Warram JH, Krolewski AS. Regression in microalbuminuria in type 1 diabetes. *New England Journal of Medicine*. 2003; 348:2285–93. [PubMed: 12788992]
- Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 2000; 56(3):779–788. [PubMed: 10985216]
- Turnbull B. The empirical distribution function with arbitrary grouped, censored, and truncated data. *Journal of the Royal Statistical Society, Series B*. 1976:290–295.
- Vonesh EF, Greene T, Schluchter MD. Shared parameter models for the joint analysis of longitudinal data and event times. *Stat in Med*. 2006; 25:143–163. [PubMed: 16025541]
- Wulfsohn M, Tsiatis A. A joint model for survival and longitudinal data measured with error. *Biometrics*. 1997; 53:330339.

Figure 1.
Complete and Missing Visit Patterns For A1c and proteinuria measurements.

Figure 2.
Relationship between level of A1c and occurrence of proteinuria in diabetic patients. Arrow indicates event