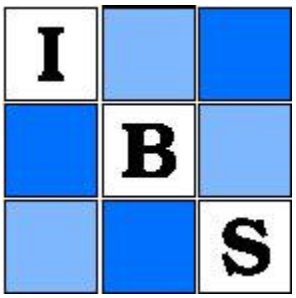

BIOMETRICS 54, 1445–1462
December 1998

Evaluating Surrogate Markers of Clinical Outcome When Measured with Error

Urania G. Dafni* **and Anastasios A. Tsiatis**

Department of Biostatistics, Harvard School of Public Health,
677 Huntington Avenue, Boston, Massachussetts 02115, U.S.A.

SUMMARY

In most clinical trials, markers are measured periodically with error. In the presence of measurement error, the naive method of using the observed marker values in the Cox model to evaluate the relationship between the marker and clinical outcome can produce biased estimates and lead to incorrect conclusions when evaluating a potential surrogate. We propose a two-stage approach to account for the measurement error and reduce the bias of the estimate. In the first stage, an empirical Bayes estimate of the time-dependent covariate is computed at each event time. In the second stage, these estimates are imputed in the Cox proportional hazards model to estimate the regression parameter of interest. We demonstrate through extensive simulations that this methodology reduces the bias of the regression estimate and correctly identifies good surrogate markers more often than the naive approach. An application evaluating CD4 count as a surrogate of disease progression in an AIDS clinical trial is presented.

1. Introduction

Randomized clinical trials are designed to evaluate new treatments efficiently. To accomplish this, patients need to be followed for long enough periods of time to reach an adequate number of endpoints. This leads to the design of either very long or very large studies when events are rare. The former can delay the determination of the outcome, and both can be prohibitively expensive. The delay in the evaluation of treatments has a big effect on the lives of patients, especially for diseases for which adequate therapy does not currently exist and the standard treatment merely prolongs survival but does not cure the disease. The AIDS epidemic spurred an increased interest directed toward using surrogate markers as a basis for treatment evaluation (Lagakos and Hoth, 1992). The potential value of surrogate markers in new drug development is enormous as new drug development can be made considerably more cost effective. It is of ethical importance to make a successful drug available to the patients whose well-being depends on it as soon as possible. On the other hand, mislabeling a marker as a good surrogate could lead to the recommendation of a noneffective drug, directing patients away from alternatives that might be more therapeutic. For these reasons, it is essential to identify a marker correctly as a good or bad surrogate of clinical outcome.

A number of requirements must be satisfied by a good surrogate marker. The marker should respond rapidly to treatment relative to the disease progression. The marker's response should imply a benefit regarding clinical outcome. A marker that is a good predictor for clinical outcome cannot necessarily be used as a surrogate for clinical outcome. According to Prentice (1989), two empirical conditions must exist to satisfy the formal definition of a surrogate marker. First, the risk of clinical outcome should correlate with the changes in the marker values over time. Second, the beneficial effect of a treatment on clinical outcome should be mediated entirely through its effect on the surrogate marker; i.e., no residual effect of treatment on the risk of clinical outcome

* *Corresponding author's email address:* dafni@sdac.harvard.edu

Key words: AIDS; Measurement error; Proportional hazards; Surrogate markers; Time-varying covariates.

should exist (Lagakos and Hoth, 1992). This implies that after adjusting for the effect of treatment on the marker, the association between the time-varying covariate and clinical outcome should not differ among various treatment groups. In summary, to show statistically that the second empirical condition is satisfied, the following should happen: The treatment effect should be significant when considered by itself in the Cox proportional hazards model; the treatment should have an effect on the marker value; and when treatment and surrogate marker are jointly included in the model, the treatment effect should disappear (i.e., be statistically nonsignificant) with an estimate of the effect close to zero.

Knowledge about the full trajectory of the marker is needed to correctly estimate the association between a time-varying marker and clinical outcome. The design and conduct of most clinical trials is such that this information is not fully available. The marker is usually known only in periodic instants in time, according to a preset time schedule defined by the study protocol. In addition, the marker is measured with some measurement error, further limiting the needed information. Even on scheduled times, information on the marker is not available for certain individuals. Finally, censoring contributes to the imperfect information.

A naive approach is to use the observed marker trajectory though it be incomplete and imperfect. This approach produces biased estimates of the association we want to estimate. Measurement error was shown by Prentice (1982) to bias the estimate toward the null. This bias affects the correct evaluation of the first empirical condition that a surrogate marker needs to satisfy. Even if the first condition is satisfied on the basis of the imperfect information (i.e., the marker is correctly identified as a good predictor of clinical outcome), the second empirical condition might not be. The imperfect information on the marker trajectory could leave a residual effect that should have been captured by the marker. This residual effect can be large enough to inflate the treatment effect to significance. The result is that using the naive technique for evaluating a marker as a potential surrogate can lead to erroneous conclusions when testing either of the two empirical conditions. If the information about the marker is imperfect enough, a potential surrogate can be incorrectly labeled unsatisfactory, with obvious consequences.

We propose a method that addresses the limitations on our knowledge of the true trajectory of the marker. A model, usually data driven, is assumed for the progression of the time-varying marker. This model relates the observed marker value with the underlying true value. Using this model, estimates of the value of the covariate are calculated at each time point at which an event occurs based on the observed history of the covariate among individuals who did not have an event up to that time point. These estimates are empirical Bayes estimates (Laird and Ware, 1982) and are then used in the Cox proportional hazards regression model. Imputing the estimates of the marker in the Cox model reduces the bias and mean squared error of the estimate of association between the time-varying covariate and survival. Without loss of generality, we explore the method in the framework of the AIDS clinical trials setting.

Our model is presented in Section 2. Sections 2.1 and 2.2 present the model for the survival relationship to the marker and treatment. Section 2.3 presents the model for the marker time progression. Section 3 shows extensive numerical results comparing the proposed method to the naive approach when interested in the first empirical condition (i.e., when interested in estimating the association of the time-varying marker to the clinical outcome). Section 4 shows numerical results regarding the second empirical condition. Section 5 presents an application of the method in evaluating the value of CD4 count as a surrogate of disease progression in a cohort of mildly symptomatic HIV-infected individuals who participated in an AIDS Clinical Trials Group protocol from 1987 to 1989 (Fischl et al., 1990). Finally, Section 6 discusses the implications of our results.

2. Model

2.1 *Modeling the Survival Relationship to the Marker and Treatment*

We want to estimate the relationship of the treatment to survival adjusted for the value of the marker. If the marker is a good surrogate, no significant relationship should exist between assigned treatment and survival. The hazard relationship to the marker should be irrelevant to the treatment an individual is on. The methodology is presented for the full model, including both the marker and the treatment in the Cox proportional hazards relationship. For ease of presentation, we assume only two treatments: a placebo and a treatment with drug A. Generalization to more than two treatments can be easily achieved.

Let $\overline{y_i(t)}$ denote the true history of the marker up to time t, T_i the survival time, and trt_i the treatment indicator for the ith individual ($i = 1, \ldots, n$); i.e., trt_i equals one if the patient is treated with drug A and zero otherwise. In most clinical trials, individuals are subject to right censoring. If we denote by C_i the potential censoring time for the ith individual, then we actually observe

$X_i = \min(T_i, C_i)$ and the corresponding failure time indicator Δ_i, where Δ_i equals one if $T_i \leq C_i$ and zero otherwise. The cause-specific hazard is given by the expression

$$\lambda^* \left(t \mid \overline{y_i(t)}, trt_i\right) = \lim_{h \to 0} \frac{P\left[X_i \in [t, t+h], \Delta_i = 1 \mid X_i \geq t, \overline{y_i(t)}, trt_i\right]}{h}. \tag{1}$$

Under the assumption of noninformative censoring, the cause-specific hazard at time t, $\lambda^*(t \mid \overline{y_i(t)}, trt_i)$, is equal to the hazard of interest, $\lambda(t \mid \overline{y_i(t)}, trt_i)$. To model the survival relationship, a Cox proportional hazards model is assumed (Cox, 1972):

$$\lambda \left(t \mid \overline{y_i(t)}, trt_i\right) = \lambda_0(t) g \left(\overline{y_i(t)}, \boldsymbol{\beta}, trt_i\right) = \lambda_0(t) g_1 \left(\overline{y_i(t)}, \beta_0\right) g_2(trt_i, \beta_1), \tag{2}$$

where $g_2(trt_i, \beta_1) = 1$ for $\beta_1 = 0$.

To estimate the parameter $\boldsymbol{\beta} = [\beta_0, \beta_1]'$, which characterizes the adjusted relationship of the marker and of the treatment to survival, one needs to maximize the partial likelihood (Cox, 1975):

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\prod_{l \epsilon D(t_i)} g \left(\overline{y_l(t_i)}, \boldsymbol{\beta}, trt_l\right)}{\left[\sum_{l \epsilon R(t_i)} g \left(\overline{y_l(t_i)}, \boldsymbol{\beta}, trt_l\right)\right]^{J_i}}, \tag{3}$$

where $i = 1, \ldots, k$ indexes $t_1 < t_2 < \cdots < t_k$ ordered distinct failure times, J_i is the number of failures at t_i, $R(t_i)$ is the risk set just prior to t_i, and $D(t_i)$ is the set of subjects failing at t_i.

Assuming that measurement error is just random noise, the observed value of the marker for individual i at time t, $z_i(t)$, can be expressed according to the following model:

$$z_i(t) = y_i(t) + \epsilon_i(t), \tag{4}$$

where $\mathrm{E}(\epsilon_i(t)) = 0$, $\mathrm{var}(\epsilon_i(t)) = \sigma^2$ and $\mathrm{cov}(\epsilon_i(t), \epsilon_i(s)) = 0$ for $s \neq t$.

The hazard is a function of the true trajectory of the marker up to time t for the ith individual. Because the true trajectory of the marker is not known and only the observed history $(z_i^H(t))$ is available, the following assumption is needed: If the true marker trajectory up to time t is known, additional knowledge of the observed history is irrelevant in determining the hazard at time t (i.e., $\lambda(t \mid \overline{y_i(t)}, z_i^H(t), trt_i)$ is equal to $\lambda(t \mid \overline{y_i(t)}, trt_i)$). Therefore, among individuals at risk at time t, neither measurement error nor the timing of observation of the marker prior to time t is prognostic.

Because the true marker trajectory is not known, a naive practice is to model the hazard directly using the observed marker values up to time t (Pepe, Self, and Prentice, 1989; Hughes, 1993). The same functional form is used on the basis of the observed history of the marker instead of the true marker trajectory with corresponding partial likelihood:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\prod_{l \epsilon D(t_i)} g \left(z_l^H(t_i), \boldsymbol{\beta}, trt_l\right)}{\left[\sum_{l \epsilon R(t_i)} g \left(z_l^H(t_i), \boldsymbol{\beta}, trt_l\right)\right]^{J_i}}. \tag{5}$$

The resulting estimate, $\hat{\boldsymbol{\beta}}$, does not necessarily converge to the parameter of interest, $\boldsymbol{\beta}$, which describes the hazard relationship to the true marker trajectory. This could lead to biased estimates of the true hazard relationship (Prentice, 1982). The actual functional hazard relationship to use in the Cox model is derived as in Pepe et al. (1989) and Tsiatis, Degruttola, and Wulfsohn (1995).

Through a simple application of the law of conditional probability, the hazard at time t, based on the history of the observed covariate, $\lambda(t \mid z_i^H(t), trt_i)$, can be expressed as the conditional expectation of $\lambda(t \mid \overline{y_i(t)}, trt_i)$ given $X_i \geq t$, trt_i, and $z_i^H(t)$:

$$\lambda \left(t \mid z_i^H(t), trt_i\right) = \lambda_0(t) \mathrm{E}_{X_i \geq t, z_i^H(t), trt_i} g(\overline{y_i(t)}, \boldsymbol{\beta}, trt_i). \tag{6}$$

For ease of presentation, it is assumed that the hazard is a function of the most recent value of the marker, $y_i(t)$, which is the biologically natural assumption for a potential surrogate marker. The arguments would be no more difficult had this relationship been a more complex functional of the history of the marker. The corresponding partial likelihood is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\prod_{l \epsilon D(t_i)} \mathrm{E}_{X_i \geq t_i, z_l^H(t_i), trt_l} [g(y_l(t_i), \boldsymbol{\beta}, trt_l)]}{\left\{\sum_{l \epsilon R(t_i)} \mathrm{E}_{X_i \geq t_i, z_l^H(t_i), trt_l} [g(y_l(t_i), \boldsymbol{\beta}, trt_l)]\right\}^{J_i}}. \tag{7}$$

Similarly to Pepe et al. (1989), we propose maximizing this likelihood to estimate the parameter of interest, $\boldsymbol{\beta}$, which describes the hazard relationship to the true value of the marker and the treatment assignment. In general, the above conditional expectation is not tractable but, for specific interesting cases, can be easily evaluated. Methods for approximating the conditional expectation are presented in the next section.

2.2 *Approximating the Conditional Expectation*

For the case of the original Cox model hazard, $\lambda(t \mid \overline{y_i(t)}, trt_i) = \lambda_0(t)e^{\beta_0 y_i(t)+\beta_1 trt_i}$; substituting in (6) for the expectation, we have

$$\lambda\left(t \mid z_i^H(t), trt_i\right) = \lambda_0(t)e^{\beta_1 trt_i} \int_{-\infty}^{\infty} e^{\beta_0 y_i(t)} f\left(y_i(t) \mid z_i^H(t), X_i \geq t, trt_i\right) d[y_i(t)]. \tag{8}$$

Note that this integral is the moment-generating function $\psi(\beta)$ for the conditional distribution of $y_i(t)$ given $z_i^H(t)$ and treatment among individuals who survived up to time t.

More specifically, if one assumes a normal conditional distribution, the partial likelihood function (7) becomes (Pepe et al., 1989):

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\prod_{l \epsilon D(t_i)} e^{\beta_1 trt_l} e^{\beta_0 \mu\left(t_i \mid z_l^H(t_i), trt_l\right) + \frac{1}{2}\beta_0^2 \sigma^2\left(t_i \mid z_l^H(t_i), trt_l\right)}}{\left[\sum_{l \epsilon R(t_i)} e^{\beta_1 trt_l} e^{\beta_0 \mu\left(t_i \mid z_l^H(t_i), trt_l\right) + \frac{1}{2}\beta_0^2 \sigma^2\left(t_i \mid z_l^H(t_i), trt_l\right)}\right]^{J_i}}, \tag{9}$$

where $\mu(t_i \mid z_l^H(t_i), trt_l)$ and $\sigma^2(t_i \mid z_l^H(t_i), trt_l)$ are the conditional mean and variance of the true value of the marker for individual l at time t_i based on the observed history of the marker up to time t_i among individuals on treatment trt_l at risk at time t_i. We use estimates $\hat{\mu}(t_i \mid z_l^H(t_i), trt_l)$ and $\hat{\sigma}^2(t_i \mid z_l^H(t_i), trt_l)$ to substitute in (9) to yield the partial likelihood estimate $\hat{L}(\beta)$. These estimates are described in more detail in the next section. As shown in Section 3, the effect of using estimates for the conditional means and variances instead of the true values is minimal.

For models in which simple expressions for $\mathrm{E}_{X_i \geq t, z_i^H(t), trt_i}[g(y_i(t), \boldsymbol{\beta}, trt_i)]$ cannot be obtained, the simple first-order approximation to the conditional expectation, also called the regression calibration approximation (Carroll, Ruppert, and Stefanski, 1995), can be used:

$$\mathrm{E}_{X_i \geq t, z_i^H(t), trt_i}[g(y_i(t), \boldsymbol{\beta}, trt_i)] = g(\mathrm{E}_{X_i \geq t, z_i^H(t), trt_i}[y_i(t)], \boldsymbol{\beta}, trt_i). \tag{10}$$

We study this method for the Cox model in great detail in later sections.

The history of the covariate up to time t for individual i, given that the individual is still at risk at time t ($X_i \geq t$), can be described as a realization of a stochastic process. In the case of a normal process and the original Cox model, the simple first-order approximation to the conditional expectation reduces the partial likelihood to

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{k} \frac{\prod_{l \epsilon D(t_i)} e^{\beta_1 trt_l} e^{\beta_0 \mu\left(t_i \mid z_l^H(t_i), trt_l\right)}}{\left[\sum_{l \epsilon R(t_i)} e^{\beta_1 trt_l} e^{\beta_0 \mu\left(t_i \mid z_l^H(t_i), trt_l\right)}\right]^{J_i}}. \tag{11}$$

This form of the partial likelihood is especially useful because the standard software for the Cox model can be used to estimate $\boldsymbol{\beta}$.

The method described here for the proportional hazards model can be easily extended to other hazard models, such as the additive relative risk model of Prentice and Self (1983).

2.3 *Modeling the Marker Time Progression*

In the AIDS clinical trials setting, the most popular candidate for a surrogate marker is an immunological marker, the CD4-lymphocyte count (Lagakos and Hoth, 1992). The pattern of CD4 count progression has been estimated in previous work (DeGruttola, Lange, and Dafni, 1991; Vittinghoff, Malani, and Jewell, 1994). The square-root or the logarithmic transformation are used to normalize the data and stabilize the variance across t. The progression of the logarithm of CD4 for HIV-infected and untreated individuals can be modeled as a linear decline with time (Tsiatis et al., 1995). The observed CD4 count is measured with error. It is assumed that the measurement error is normally distributed with mean zero and variance σ_ϵ^2 on the logarithmic scale. One could envision a successful treatment capable of significantly increasing survival while linearly increasing the logarithm of CD4 with time. In this scenario, one could test whether CD4 count is a good surrogate marker.

In this section, we model the joint relationship of $y_i(t)$ and $z_i^H(t)$ as a normal stochastic process. We concentrate on a time-varying marker, a transformation of which, $y_i(t)$, follows a normal distribution and its progression is linear with time. One could easily use higher-order models, such as a quadratic model. These models offer a great deal of flexibility. For individual i $(i = 1, \ldots, n)$ receiving treatment k at time t, the linear growth curve model is

$$y_i(t) = \alpha_{0_{ik}} + \alpha_{1_{ik}} t, \tag{12}$$

where $(\alpha_{0_{ik}}, \alpha_{1_{ik}})' \sim \text{BVN}(\boldsymbol{\theta}_k, \boldsymbol{\Theta}_k)$, $\boldsymbol{\theta}_k' = (\theta_{0k}, \theta_{1k})'$ and

$$\boldsymbol{\Theta}_k = \begin{bmatrix} \sigma_{0k}^2 & \sigma_{01k} \\ \sigma_{01k} & \sigma_{1k}^2 \end{bmatrix}.$$

The marker is measured with error. The assumption is made that measurement error is normally distributed in the transformed scale with mean zero and variance σ_ϵ^2. The model for the normalizing transformation of the observed marker, $z_i(t)$, is

$$z_i(t) = \alpha_{0_{ik}} + \alpha_{1_{ik}} t + \epsilon_i, \tag{13}$$

where $\epsilon_i \sim \text{N}(0, \sigma_\epsilon^2)$, $(\alpha_{0_{ik}}, \alpha_{1_{ik}})' \sim \text{BVN}(\boldsymbol{\theta}_k, \boldsymbol{\Theta}_k)$, and ϵ_i, $(\alpha_{0_{ik}}, \alpha_{1_{ik}})'$ independent.

At any time t, for individuals who survived up to that time, a history of observation values for the marker is available, $z_i^H(t) = (z_i(t_{i1}), z_i(t_{i2}), \ldots, z_i(t_{im}))$, where $j = 1, \ldots, m$ indexes the observed times up to time t. On the basis of that history, the true value of the surrogate marker can be estimated at time t for individual i in treatment k by the conditional expectation

$$\text{E}_{T \geq t, trt_i = k}(y_i(t) \mid (z_i(t_{i1}), \ldots, z_i(t_{im})), trt_i = k).$$

According to (12) and (13) the joint distribution of $y_i(t)$ and $z_i^H(t)$ is multivariate normal with mean vector μ_k and variance–covariance matrix $\Sigma^{(k)}$:

$$\begin{bmatrix} y_i(t) \\ z_i(t_{i1}) \\ z_i(t_{i2}) \\ \vdots \\ z_i(t_{im}) \end{bmatrix} \sim \text{N}\left(\begin{bmatrix} \theta_{0k} + \theta_{1k} t \\ \theta_{0k} + \theta_{1k} t_{i1} \\ \theta_{0k} + \theta_{1k} t_{i2} \\ \vdots \\ \theta_{0k} + \theta_{1k} t_{im} \end{bmatrix}, \Sigma^{(k)} \right). \tag{14}$$

Partitioning the $(m+1) \times (m+1)$ matrix $\Sigma^{(k)}$, we have

$$\Sigma^{(k)} = \begin{bmatrix} \Sigma_{11}^{(k)} & \Sigma_{12}^{(k)} \\ \Sigma_{21}^{(k)} & \Sigma_{22}^{(k)} \end{bmatrix}, \tag{15}$$

where

$$\Sigma_{11}^{(k)} = \sigma_{0k}^2 + 2t\sigma_{01k} + t^2\sigma_{1k}^2,$$
$$\Sigma_{12}^{(k)} = \Sigma_{21}^{(k)'} = \left[\sigma_{0k}^2 + (t + t_{i1})\sigma_{01k} + tt_{i1}\sigma_{1k}^2, \ldots, \sigma_{0k}^2 + (t + t_{im})\sigma_{01k} + tt_{im}\sigma_{1k}^2\right],$$

and

$$\Sigma_{22}^{(k)} = \begin{bmatrix} [1 \quad t_{i1}]\boldsymbol{\Theta}_k \begin{bmatrix} 1 \\ t_{i1} \end{bmatrix} + \sigma_\epsilon^2 & \cdots & [1 \quad t_{i1}]\boldsymbol{\Theta}_k \begin{bmatrix} 1 \\ t_{im} \end{bmatrix} \\ \vdots & \ddots & \\ [1 \quad t_{im}]\boldsymbol{\Theta}_k \begin{bmatrix} 1 \\ t_{i1} \end{bmatrix} & \cdots & [1 \quad t_{im}]\boldsymbol{\Theta}_k \begin{bmatrix} 1 \\ t_{im} \end{bmatrix} + \sigma_\epsilon^2 \end{bmatrix}.$$

From well-known results about normal distributions (Guttman, 1982), we have that the conditional distribution of $y_i(t)$, given the history of the observed covariate and treatment k, is normal with mean $\mu_k(t \mid z_i^H(t))$ and variance $\sigma_k^2(t \mid z_i^H(t))$ given by

$$\mu_k\left(t \mid z_i^H(t)\right) = \theta_{0k} + \theta_{1k} t - C_{11}^{(k)^{-1}} C_{12}^{(k)} \left(\begin{bmatrix} z_i(t_{i1}) \\ z_i(t_{i2}) \\ \vdots \\ z_i(t_{im}) \end{bmatrix} - \begin{bmatrix} \theta_{0k} + \theta_{1k} t_{i1} \\ \theta_{0k} + \theta_{1k} t_{i2} \\ \vdots \\ \theta_{0k} + \theta_{1k} t_{im} \end{bmatrix} \right) \tag{16}$$

and $\sigma_k^2(t \mid z_i^H(t)) = C_{11}^{(k)^{-1}}$ with $C_{11}^{(k)} = (\Sigma_{11}^{(k)} - \Sigma_{12}^{(k)}\Sigma_{22}^{(k)^{-1}}\Sigma_{21}^{(k)})^{-1}$ and $C_{12}^{(k)} = (\Sigma_{12}^{(k)}\Sigma_{22}^{(k)^{-1}}\Sigma_{21}^{(k)} - \Sigma_{11}^{(k)})^{-1}\Sigma_{12}^{(k)}\Sigma_{22}^{(k)^{-1}}$. These conditional means $\mu_k(t \mid z_i^H(t))$ and variances $\sigma_k^2(t \mid z_i^H(t))$ for individual i at time t are used in the partial likelihood given in equation (9). The conditional variance $\sigma^2(t_i \mid z_l^H(t_i), trt_i)$ is a function of the vector of observation times up to time t_i and Θ_k. If we assume common variance–covariance matrix Θ across treatments and deal with the complete case (i.e., no missingness on scheduled times) with the same schedule of observing the marker among all individuals, the term involving the conditional variance drops from the partial likelihood given in equation (9), reducing to the partial likelihood given in equation (11).

The method used to estimate the parameters of the normal distribution (θ, Θ, σ_ϵ^2) is presented in the Appendix and is similar to a method presented by Jones and Ackerson (1990).

3. First Empirical Condition: Numerical Results

Simulations were used to evaluate the new methodology. We compare our proposed method to the naive approach, when the primary interest lies in estimating the association of the time-varying covariate to clinical outcome. Each simulation run was equivalent to a study following 100 untreated HIV-infected individuals. The time-varying covariate here is the value of the logarithm of the CD4 counts at each time point. As discussed earlier, the CD4 count is the most popular candidate for a surrogate marker in the AIDS clinical trials setting (Lagakos and Hoth, 1992), and the progression of the logarithm of CD4 for HIV-infected individuals can be modeled as a linear decline with time (Tsiatis et al., 1995). We can directly apply the described methodology for the normal process, linear growth curve model, and original Cox hazard model to compare numerically our method to the naive approach. Each simulated study was chosen to last for a 72-week period with looks at the surrogate marker at intervals of 8 weeks. The censoring distribution was assumed to be exponential ($\lambda = 0.00909$) with a probability of a patient being followed through week 72 of 0.52. Additional censoring occurred at the end of the study. The survival data were created with a Cox model beta of -1.0 and a baseline hazard of 1.0 constant across time. For each set of conditions, 50 simulated studies were analyzed.

The initial data we used mimicked the data from the placebo arm of the Burroughs–Wellcome 02 (BW-02) study. The values of the parameters θ_0, θ_1; their variance–covariance matrix Θ; and the error variance σ_ϵ^2 were taken to equal the restricted maximum likelihood estimates (reml) at week 4 for the placebo arm of the BW-02 study. These were

$$\begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix} = \begin{bmatrix} 1.24 & -0.0114 \\ -0.0114 & 0.003 \end{bmatrix},$$

where error variance $\sigma_\epsilon^2 = 0.32$. Furthermore, the mean CD4 at baseline was around 65 counts, which corresponds to an intercept of 4.173 on the log scale (θ_0). A negative slope of -0.0103 on the log scale was used to describe the linear decline of the transformed CD4 (θ_1). With these parameters, according to the model, at week 72 the mean CD4 will be around 30 counts (3.432 on the log scale).

Results for the complete case are presented in Table 1. In the 50 simulations, the number of events in the 72 weeks varied from 46% to 66%, with a median of 57.5%.

In the ideal situation, complete knowledge of the trajectory of the CD4 counts over time is available, no experimental error exists ($\sigma_\epsilon^2 = 0$), and CD4 is known exactly at each event time. Summary statistics for the estimates from 50 simulated data sets using the true value at the time of an event for the time-varying covariate in the Cox model are presented in the first row of Table 1.

In a less ideal situation, but one that is closer to reality, CD4 counts are measured only periodically. We still assume that measurement error is nonexistent and that the true CD4 counts are available in intervals of 8 weeks for the duration of the study. The CD4 value measured just prior to the event time is then used in the Cox regression model. We see that the resulting estimate of association of survival to the CD4 counts is slightly biased but still very close to the true β value of -1.0 (Table 1, second row).

In reality, CD4 counts are observed periodically and with a substantial amount of variability arising from measurement error. We use three different values for σ_ϵ^2—0.32, 0.62, and 1.24—which correspond to measurement error contributing about 20%, 33%, and 50%, respectively, of the total variability of the baseline CD4 count in the transformed scale. According to the proposed method, the intercept θ_0, the slope θ_1, the variance–covariance matrix Θ, and the error variance σ_ϵ^2 are estimated in intervals of 8 weeks using the history of the time-varying covariate for all individuals

Table 1
Complete data; summary statistics for the estimate of association of survival to time-varying log CD4

Beta estimate	Mean	SD	Bias	MSE
True log CD4 at event times	−0.99873	0.11474	0.00127	0.013168
True log CD4 every 8 weeks	−1.04690	0.11709	−0.04690	0.015910
$\sigma_\epsilon^2 = 0.32$				
Estimated log CD4	−0.95081	0.12139	0.04919	0.017156
Observed log CD4	−0.81466	0.10277	0.18534	0.044912
$\sigma_\epsilon^2 = 0.62$				
Estimated log CD4	--0.92168	0.13063	0.07832	0.023196
Observed log CD4	−0.68017	0.09510	0.31983	0.111335
$\sigma_\epsilon^2 = 1.24$				
Estimated log CD4	−0.87903	0.15283	0.12097	0.037991
Observed log CD4	−0.51147	0.08301	0.48853	0.245548

at risk at the beginning of the 8-week period. The logarithm of CD4 is estimated according to the linear time progression model at each event time. These values are then imputed in the Cox model. Table 1 presents a comparison of the estimates for each different value of σ_ϵ^2 when imputing the estimated log CD4 counts in the Cox regression model according to the proposed method and when simply using the observed values.

For $\sigma_\epsilon^2 = 0.32$, the 95% confidence intervals for β when imputing the estimated values for the covariate in the Cox model covered the true value for 46 of the 50 simulations (coverage = 92.0% of the time) and the Cox model with observed covariates only for 34 simulations (coverage = 68.0%). Both estimates are biased toward the null. Using the estimated CD4s instead of the observed gives an improved estimate of beta, with less bias and smaller mean squared error. In fact, our estimate for the case of $\sigma_\epsilon^2 = 1.24$ is less biased than the estimate based on the naive approach when $\sigma_\epsilon^2 = 0.32$. The corresponding coverage probabilities for $\sigma_\epsilon^2 = 0.62$ and $\sigma_\epsilon^2 = 1.24$ based on our model were 90% and 80%, whereas those based on the naive approach were 14% and 0%, respectively.

In most clinical trials, patients do not come to all visits. The situation of having a certain percentage of the time-varying covariates missing was considered. The initial data conditions were used for 50 simulations. Observed log CD4 counts were randomly missing with probability 25%. The simulation results are presented in Table 2.

As is shown, the 25% level of random missingness has no significant effect on the estimate. The results are similar to the ones presented in Table 1. It is of interest to note that using the first-order approximation yields almost identical estimates to using the moment-generating function in the Cox model. For this and other patterns of missingness considered, either method is preferable to the naive approach.

The effect of using estimates for the unknown components of the induced relative risks on the estimate of β is explored by the direct comparison of two different estimates of its variance (Table 3).

The first estimate (Var1) is the square of the standard deviation of the 50 estimates of β produced by the 50 simulations; the second estimate (Var2) is the average of the 50 variance estimates produced by the maximization of the partial likelihood in the 50 simulations. As can be seen from Table 3, the two estimates are very close to each other; i.e., using estimates in the partial likelihood has minimal effect on the variance of β.

Table 2
Missing data; summary statistics for the estimate of association of survival to time-varying log CD4 / $\sigma_\epsilon^2 = 0.32$ / $p = 0.25$

Beta estimate	Mean	SD	Bias	MSE
First-order approximation	−0.95373	0.11754	0.04627	0.015957
Moment-generating function	−0.94238	0.12422	0.05763	0.018752
Observed log CD4	−0.81894	0.10453	0.18106	0.043708

Table 3
Estimates for the variance of β

Variance Estimate	Var1	Var2
True log CD4 at event times	0.0131660	0.0172115
True log CD4 every 8 weeks	0.0137103	0.0192790
$\sigma_\epsilon^2 = 0.32$		
Estimated log CD4	0.0147364	0.0181618
Observed log CD4	0.0105609	0.0136910
$\sigma_\epsilon^2 = 0.62$		
Estimated log CD4	0.0170629	0.0187193
Observed log CD4	0.0090437	0.0108824
$\sigma_\epsilon^2 = 1.24$		
Estimated log CD4	0.0233561	0.0199059
Observed log CD4	0.0068898	0.0077214

3.1 *Discussion*

Assuming that one models the marker progression and the hazard function correctly, our proposed method produces an estimate with much less bias and smaller mean squared error than when simply using the observed covariates in the Cox model. One of the implicit assumptions in our technique is that, at each risk set, the distribution of the observed covariate is normal. In reality, individuals with lower-valued covariates would be expected to die earlier so that fewer of the covariate values on the lower end of the distribution will be included in the later risk sets. This might be an explanation for the bias of the estimate not totally disappearing with our model. A proposed remedy of the problem might be the assumption of a more complex covariate distribution function that is capable of capturing the skewness in the data. Davidian and Gallant (1992) proposed an appropriate family of distributions.

Alternatively, one might want to incorporate information on how covariates are dropping from the risk set in modeling their distribution at each event time. An iterative procedure could be used for modeling the joint distribution of α_{0i}, α_{1i} given $z_i^H(t)$ and $X_i \geq t$. Then the conditional expectation of the hazard function among those at risk in (6) can be estimated using the joint conditional distribution $P(\alpha_{0i}, \alpha_{1i} \mid z_i^H(t), X_i \geq t)$, which incorporates the actual hazard. The association of survival to the time-varying covariates, β, is estimated by maximizing this partial likelihood.

4. Second Empirical Condition: Numerical Results

Next, simulations were used to compare our proposed method to the naive approach, when our primary interest is to identify a good surrogate marker. For a marker to be a good surrogate, no significant relationship should exist between assigned treatment and clinical outcome in the presence of the marker in the Cox model. Each simulation run represented a placebo-controlled clinical trial on a total of 200 patients. The simulated study was chosen, similarly to before, to last for a 72-week period with looks at the surrogate marker at intervals of 8 weeks. One hundred patients were followed on each of the two treatment arms. The censoring distribution, common for both arms, was created as presented in Section 3.

In the placebo arm, the surrogate marker progression was represented by a linear decline in time. In fact, the placebo arm of the study was taken to be identical to the study described in section 3.

We assumed that therapy was successful in substantially increasing the CD4 counts of treated individuals. Thus, in the treatment arm, the surrogate marker progression was represented by a linear growth curve model with a positive slope on the logarithmic scale. The slope was taken to be 0.0412 (θ_1), which, starting with the same baseline as for the placebo patients (around 65 counts, or an intercept $\theta_0 = 4.173$ on the log scale), results to a mean CD4 at week 72 of 1260 counts (7.139 on the log scale). All other parameters—σ_0^2, σ_{01}^2, σ_1^2, and σ_ϵ^2—were set the same for both treatments (as in Section 3).

The survival distribution was created by the original Cox hazard model with $\beta_0 = -1.0$, $\beta_1 = 0.0$ and a baseline hazard of 1.0 constant across time. According to this model, the hazard at time t for the ith individual is a function of the value of the logarithm of the CD4 counts at time t for this individual and independent of treatment. All the survival differences detected between the study arms are a direct result of the beneficial effect of the treatment on the subjects' CD4 counts.

For 41 of the 50 randomly generated studies, the difference in survival between treatments was statistically significant.

We have created the data using the premise that we have a good surrogate marker. Knowledge of a patient's treatment should not add any information in determining the patient's chances of survival in the next time instant if we already know his current value of CD4 counts. Of main interest is to identify the method that most often correctly classifies this marker as being good. Two criteria are used to identify the best method. In the presence of the CD4 counts in the Cox model, one should be able first to characterize the treatment effect as nonsignificant and second to estimate the treatment coefficient close to zero. The method that best satisfies these criteria is the preferred method. As noted earlier, for the case of a normal process and the original Cox hazard model, using either the empirical Bayes estimate in the Cox model (the first-order approximation) or the moment-generating function results in the same partial likelihood presented in equation (11).

The results from 50 simulations for complete data are presented in Table 4. The estimated treatment effect when treatment was the only covariate in the model is given in the first row of Table 4 (Model A). Summary statistics for the ideal case (i.e., CD4 is known exactly at each event time) are presented in the second and third rows of Table 4 (Model B). We see from the results that the treatment effect disappears in the presence of the true value of the surrogate marker. The 95% confidence intervals, in the presence of the true log CD4 count, covered the true value of the treatment effect (zero) in 45 of the 50 studies. We concluded, correctly, that the marker is a good surrogate 90% of the time.

Next, we assume that measurement error is nonexistent and that true CD4 counts are available in intervals of 8 weeks for the duration of the study. The CD4 value measured just prior to the event time is then used in the Cox regression model. The treatment effect estimate is not as close to zero as before, although the estimate of the association of the marker to clinical outcome is very close to the true value of $\beta_0 = -1.0$ (Table 4, Model C). From the 50 studies, using a 95% confidence interval, we concluded, correctly, that the marker is a good surrogate 88% of the time in the presence of the true log CD4 count every 8 weeks.

In reality, CD4 counts are observed periodically and with a substantial amount of variability arising from measurement error. As before, we consider three different values for σ_ϵ^2: 0.32, 0.62, and 1.24.

Table 4

Summary statistics for the estimate of association of survival to time-varying log-CD4 count and treatment

Model	Covariate	Parameter estimate	SD	Bias	MSE
A	Treatment	−0.5959390	0.236958		
B	True log CD4 at event times	−1.0201660	0.095842	−0.0201660	0.009592
	Treatment	−0.0160230	0.267347	0.0160230	0.071730
C	True log CD4 every 8 weeks	−1.0664620	0.097576	−0.0664620	0.013938
	Treatment	−0.1258590	0.266668	0.1258590	0.086952
	$\sigma_\epsilon^2 = 0.32$				
D	Estimated log CD4	−0.9679420	0.092865	0.0320580	0.009652
	Treatment	−0.0194304	0.298178	0.0194304	0.089288
E	Observed log CD4	−0.8365020	0.083089	0.1634980	0.033636
	Treatment	−0.2208706	0.263175	0.2208706	0.118045
	$\sigma_\epsilon^2 = 0.62$				
D	Estimated log CD4	−0.9358460	0.092910	0.0641540	0.012748
	Treatment	−0.0360120	0.307501	0.0360120	0.095854
E	Observed log CD4	−0.7017100	0.075927	0.2982910	0.094742
	Treatment	−0.2751987	0.256818	0.2751987	0.141690
	$\sigma_\epsilon^2 = 1.24$				
D	Estimated log CD4	−0.8882410	0.097433	0.1117590	0.021983
	Treatment	−0.0605411	0.316968	0.0605411	0.104134
E	Observed log CD4	−0.5317400	0.065018	0.2982910	0.094742
	Treatment	−0.3457457	0.248339	0.3457457	0.181212

Model D uses the empirical Bayes estimates of the log CD4 counts at the event time according to our proposed model. Model E uses the observed values of log CD4 every 8 weeks in the Cox model.

In comparing the estimates from our proposed model to the naive approach, we see that our estimates of association of clinical outcome to the marker and treatment are much less biased. For $\sigma_\epsilon^2 = 0.32$, our method is as good, if not better, at estimating the association of the marker to clinical outcome as when the true CD4 count is available every 8 weeks. The naive approach results in biased estimates for both the association of the clinical outcome to the marker ($\beta_0 = -1.0$) and the treatment ($\beta_1 = 0.0$).

In addition, the treatment effect estimate is very close to zero. For $\sigma_\epsilon^2 = 0.32$, using the 95% confidence interval, in the presence of the empirical Bayes estimate for log CD4 count, we concluded, correctly, that the marker is a good surrogate 88% of the time, the same as if we knew the true CD4 count every 8 weeks (Model C). Using the naive approach, we reached the correct conclusion only 78% of the time.

For $\sigma_\epsilon^2 = 0.62$, our method concluded, correctly, that the marker is a good surrogate 86% of the time, compared to 72% when using the naive approach, whereas for $\sigma_\epsilon^2 = 1.24$, the corresponding coverage probabilities are 86% and 62%, respectively. In the presence of larger measurement error, the bias of the estimates based on observed CD4 increases dramatically.

4.1 *Discussion*

In comparison to the naive approach using the empirical Bayes estimates in the Cox model results in much less biased estimates of association of the clinical outcome to both the marker and the treatment. When evaluating a surrogate marker according to the second empirical condition using our proposed method, the correct conclusion that the marker is a good surrogate, because all treatment effect is mediated through it, is reached with a much higher percentage. However, one should keep in mind that for these simulations the true underlying hazard and marker progression models were used for the data because it was known how the data were generated for the simulation studies.

5. Application of the Proposed Method in Evaluating log CD4 counts as a Surrogate of Disease Progression in a Cohort of HIV-Infected Individuals

5.1 *Introduction*

Data from large-scale comparative trials are required to assess whether a potential surrogate marker complies with the two empirical conditions set by Prentice (1989). A few large-scale multicenter clinical trials conducted by the AIDS Clinical Trials Group have been used to evaluate surrogate markers in HIV-infected individuals. The CD4 lymphocyte count has been evaluated for patients with AIDS as a surrogate for survival (Tsiatis et al., 1995) and for asymptomatic HIV patients as a surrogate for disease progression (Choi et al., 1993). The results consistently indicate that, although CD4 count significantly correlates with disease progression, very little of the effect of ZDV treatment on disease progression is mediated through its effect on the CD4 count.

In this paper, we evaluate the value of the CD4 count as a surrogate of disease progression in a cohort of mildly symptomatic HIV-infected individuals. The goal of the analysis presented here is to study whether CD4 count can serve as a useful surrogate for disease progression in this cohort.

Section 5.2 describes the characteristics of the study population of interest. Section 5.3 discusses the model for the relationship of the time-varying CD4 counts to clinical progression to AIDS. Section 5.4 presents the model of time progression for the CD4 counts. Section 5.5 presents the results of our analysis. Finally, Section 5.6 discusses our results.

5.2 *Study Population*

The study population consists of patients from a completed study by the AIDS Clinical Trials Group (ACTG) that evaluated the safety and efficacy of zidovudine (ZDV) in the treatment of patients with mildly symptomatic human immunodeficiency virus type 1 (HIV) infection (Fischl et al., 1990). The study was a double-blind, randomized, placebo-controlled, multicenter trial performed at AIDS clinical trials units. Following activation in August 1987, a total of 711 patients were randomized to either 200 mg ZDV taken orally every 4 hours (350 patients) or placebo (361 patients). Accrual was terminated in May 1989, and subjects were to be followed for 2 additional years. In August 1989, however, a lower rate of progression to more advanced HIV disease among ZDV-treated patients led to an independent data and safety monitoring board recommendation to terminate the study prematurely. On August 3, 1989, placebo patients were informed of their treatment code and crossed over to the ZDV treatment arm.

In our analysis, we use data only up to the crossover date. The median duration of follow-up was 50 weeks. An event was defined as death or progression to AIDS by documented onset of an opportunistic infection or Kaposi's sarcoma. There were no significant differences between the two treatment groups in pretreatment characteristics (Fischl et. al, 1990). Twenty-seven events were observed in the placebo arm during this period versus 10 in the ZDV arm.

According to the study design, several immunological measures (e.g., CD4 count, CD4 percentage, CD8 count and percentage, and ratio of CD4 to CD8) were observed at prespecified time intervals. Evaluations were scheduled every 4 weeks up to week 16, at weeks 24 and 40, and then subsequently every 12 weeks up to week 100. The CD4 data collected at these scheduled visits are used in our analysis.

5.3 *The Hazard Model*

We want to estimate the relationship of the treatment to survival, adjusted for the value of the CD4 count. For CD4 to be a good surrogate marker, no significant relationship should exist between assigned treatment and survival in the presence of CD4 in the Cox model (Prentice, 1989). The hazard relationship to CD4 should be the same for both treatment groups. We are looking at patients randomized to two treatment arms: placebo and ZDV treatment.

To model the survival relationship, a Cox proportional hazards model is assumed. Let $z_i^H(t)$ denote the observed history of the log CD4 counts up to time t, T_i the survival time, and trt_i the treatment indicator for the ith individual ($i = 1, \ldots, n$); i.e., trt_i equals one if the patient is treated with ZDV and zero otherwise.

Using the first-order approximation, the hazard at time t as a function of log CD4 count, given the observed history $z_i^H(t)$ and treatment among individuals who survived up to time t, is given by

$$\lambda\left(t \mid z_i^H(t), trt_i\right) = \lambda_0(t) e^{\beta_1 trt_i} e^{\beta_0 \mu\left(t \mid z_i^H(t), trt_i\right)}, \tag{17}$$

where $\mu(t \mid z_i^H(t), trt_i)$ is the conditional mean for the log CD4 counts at every distinct failure time given the observed history of the log CD4 counts up to that time. We want to maximize the partial likelihood function given by (11).

Extensive justification for using the estimated instead of the observed covariate value in the hazard function is given here in Sections 2 through 4. The proposed method produces an estimate of $\boldsymbol{\beta}$ with less bias and mean squared error than the standard method of using the observed value of the covariate.

5.4 *Modeling the* CD4 *Count Time Progression*

The CD4 count is the time-varying covariate of interest. A logarithmic transformation is used to normalize the CD4 data (Tsiatis et al., 1995). For patients receiving ZDV, the data suggest an initial increase in log CD4 count during the first few weeks on therapy, with a peak at around 8 weeks followed by a linear decline (Figures 1 and 2). A growth curve model with random effects is used to model the log CD4 trajectory (Laird and Ware, 1982). For untreated individuals, a linear decline best describes the trajectory of the CD4 on the logarithmic scale (Figures 1 and 2). However, the same less parsimonious model, as for ZDV treated individuals, is used to allow for similar CD4 behavior in the two treatment arms. The model parameters are treatment specific. Figure 2 shows how close the model (dotted line) fits the observed log-CD4 count trajectory (solid line).

For individual i ($i = 1, \ldots, n$) on treatment k at time t, the model for the true CD4 count progression on the logarithmic scale is

$$y_i(t) = \alpha_{0_{ik}} + \alpha_{1_{ik}} \min(t, 8) + \alpha_{2_{ik}} \max(t - 8, 0), \tag{18}$$

where $(\alpha_{0_{ik}}, \alpha_{1_{ik}}, \alpha_{2_{ik}})' \sim \text{MVN}(\boldsymbol{\theta}_k, \boldsymbol{\Theta}_k)$.

The CD4 count is measured with error. The assumption is made that measurement error is normally distributed in the transformed scale with mean zero and variance σ_ϵ^2. For individual i ($i = 1, \ldots, n$) on treatment k at time t, the model for observed CD4 on the logarithmic scale is

$$z_i(t) = \alpha_{0_{ik}} + \alpha_{1_{ik}} \min(t, 8) + \alpha_{2_{ik}} \max(t - 8, 0) + \epsilon_i,$$

where $\epsilon_i \sim \text{N}(0, \sigma_\epsilon^2)$, $(\alpha_{0_{ik}}, \alpha_{1_{ik}}, \alpha_{2_{ik}})' \sim \text{MVN}(\boldsymbol{\theta}_k, \boldsymbol{\Theta}_k)$, and ϵ_i, $(\alpha_{0_{ik}}, \alpha_{1_{ik}}, \alpha_{2_{ik}})'$ independent.

On the basis of this model, we calculate restricted maximum likelihood estimates (reml) for our parameters. The reml estimate for $\mu_k(t_i \mid z_l^H(t_i))$ is an empirical Bayes estimate (Laird and Ware, 1982).

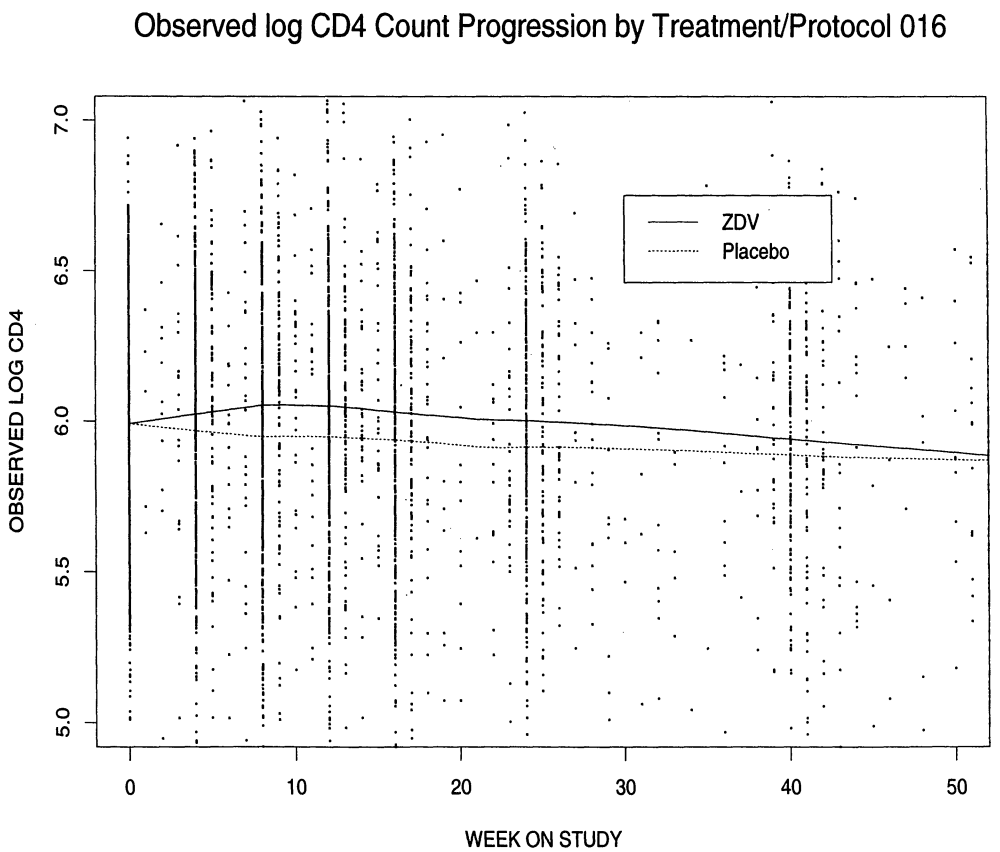

Figure 1. Observed log-CD4 count progression—Lowess curves.

According to the proposed method, the trajectory parameters $\boldsymbol{\theta}_k$, the variance–covariance matrix $\boldsymbol{\Theta}_k$, and the error variance σ_ϵ^2 are estimated at each event week using the history of the time-varying covariate for all individuals at risk on treatment k at the beginning of the event week. The logarithm of CD4 is estimated according to the time progression model at each event time. These estimates are then imputed in the Cox proportional hazards regression model. Maximizing the partial likelihood yields an estimate for $\boldsymbol{\beta}$.

5.5 *Results: Is* CD4 *a Good Surrogate Marker for Progression to AIDS?*

The estimated log-CD4 count was compared at each event time to other components of the CD4 count trajectory, i.e., baseline value (intercept), slope from baseline to week 8, peak value at week 8, and slope subsequent to week 8. These components alone and in combination were used in the Cox model with and without the presence of the estimated log-CD4 count value at each event time. Separate analyses were performed for ZDV treatment and placebo, as was a third analysis that included all randomized patients. The results from all three analyses showed that the most recent log-CD4 count was a far better predictor than any other component of the log-CD4 count trajectory or any combination thereof. No trajectory component remained significant in the presence of the estimated log-CD4 count at each event time in the Cox model.

In addition, the question whether prognosis changes if a patient's CD4 count drops below a certain level was addressed. This question is of particular importance because it would have an effect on treatment decisions and patient management. Several indicators were created to be one if a patient's CD4 count reached a specified level and zero otherwise. For example, if a patient's CD4 count drops below $300/\text{mm}^3$ counts for two consecutive visits 4 weeks apart, the corresponding indicator becomes one. Other levels of interest were 200 counts, 150, 100, 50 per mm^3, and dropping 50% below the baseline CD4 count. All these indicators were compared for surrogacy to the most recent log-CD4 count value.

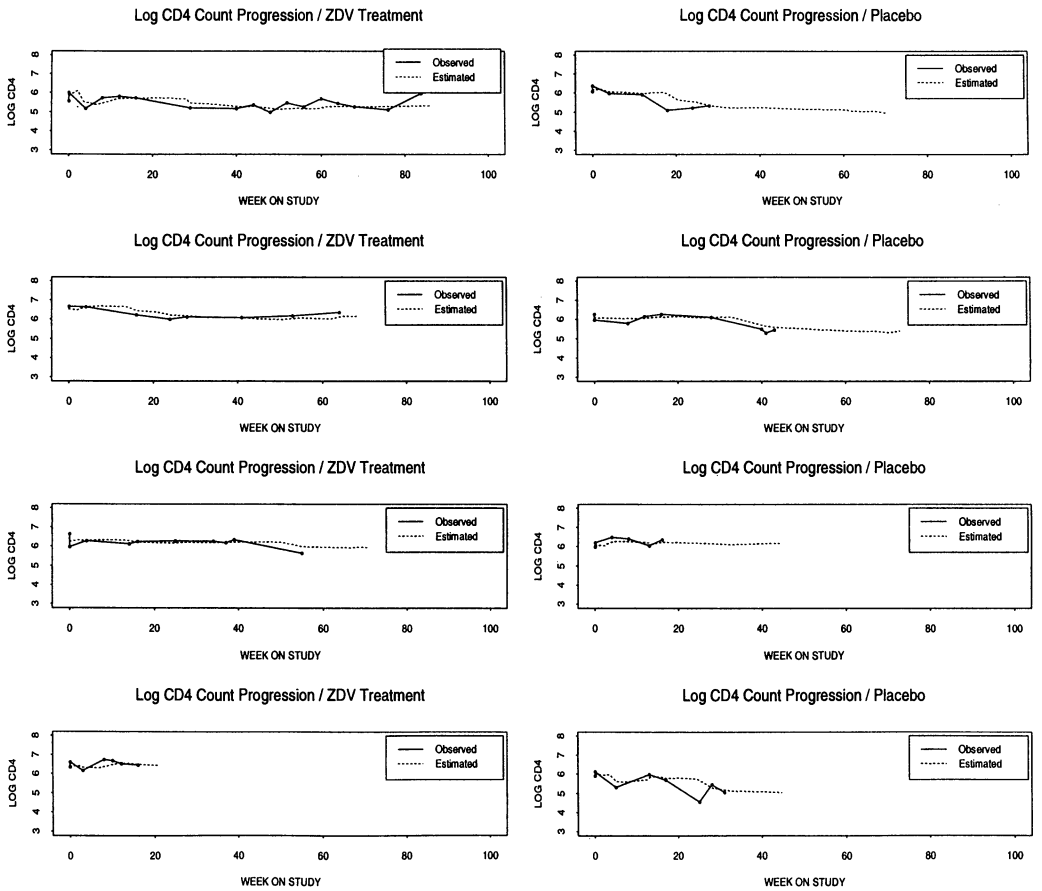

Figure 2. Observed and estimated log-CD4 count progression—first few patients.

The estimates of the most recent log-CD4 count and indicators of CD4 levels (e.g., CD4 below 300, CD4 below 200, and 50% drop) were used alone and in combination as time-varying covariates in the Cox proportional hazards regression model. Each indicator used alone in the Cox model was significant. When combined with other indicators, falling below 200 was the most significant, followed by the 50% drop indicator and then by the indicator for CD4 falling below 300. No time-varying covariate alone or in combination proved to be a good surrogate marker. Comparing the most recent log-CD4 count to the indicators separately for each of the treatment groups led to the same conclusion; i.e., recent log-CD4 count used alone in the Cox model was most predictive of disease progression. An analysis performed on patients who were still at risk at week 24 led to identical conclusions. The time point of 24 weeks was chosen to be after the initial effect of ZDV treatment on the CD4 counts; this effect appears to be transient.

To assess whether the most recent CD4 count is a good surrogate for clinical progression, we imputed the empirical Bayes estimate for log CD4 at each event time, $\mu_k(t_i \mid z_i^H(t_i))$, in the Cox model with and without the presence of the treatment indicator.

The results in the first row of Table 5, corresponding to the model with recent log-CD4 count as the only covariate, indicate that recent log-CD4 is a strong predictor of clinical outcome (Model

Table 5

Estimate of association of disease progression to time-varying log-CD4 count and treatment

Model	Covariate	Parameter estimate	SE	P value
A	Estimated log CD4	−1.737933	0.30706	0.0001
B	Treatment	−1.033366	0.37040	0.0053
C	Estimated log CD4	−1.677605	0.32509	0.0001
	Treatment	−0.828565	0.37262	0.0262

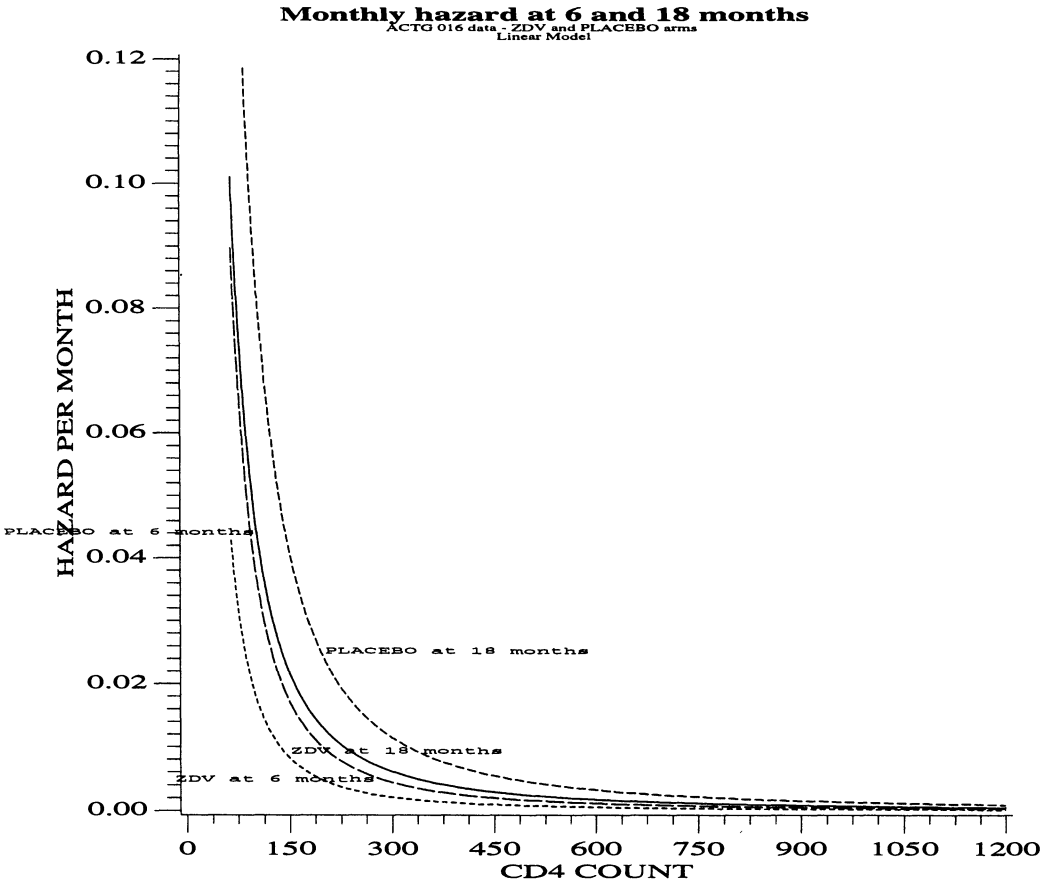

Figure 3. Monthly hazard for ZDV and placebo at 6 and 18 months.

A). The second row, corresponding to the model with treatment as the only covariate, shows a significant difference in clinical outcome between treatments (Model B). When including both most recent log-CD4 count and treatment in the Cox model, we see that the treatment effect remains significant (Model C). The treatment relative risk changes only slightly in the presence of the log-CD4 counts in the model (from 0.356 to 0.437). Thus, although the CD4 count in the logarithmic scale is a good predictor of clinical outcome, very little of the treatment effect is explained by the beneficial effect of treatment on log-CD4 count. We conclude that the CD4 count on the logarithmic scale is a good predictor of the clinical outcome but falls well short of being a good surrogate marker.

Figure 3 shows the relationship of the hazard function to CD4 counts at 6 and 18 months for placebo and ZDV patients. A smoothed Breslow estimator is used for the hazard function (Breslow, 1974). This figure shows that as CD4 declines, the hazard rate increases. The greatest effect seems to be at around a CD4 count of $200/\text{mm}^3$. In addition, a time trend indicates an increase in the hazard as a function of time even though we adjusted for the CD4 values. Finally, there is clearly a higher hazard rate for placebo patients versus ZDV-treated patients. This implies that the beneficial effect of ZDV is not mediated entirely through its effect on CD4 counts.

The same conclusion can be drawn from Figure 4. The patients who received ZDV survived far longer than would have been predicted by the ZDV effect on their CD4 counts alone. Most of the ZDV treatment effect is not explained by the effect on the CD4 count. Although CD4 counts are highly predictive of disease progression, they cannot be used as a good surrogate marker.

5.6 *Discussion*

The results presented here are consistent with the results reported in other studies casting doubt on the use of CD4 count as a good surrogate marker (Choi et al., 1993; Tsiatis et al., 1995). The CD4 count, although a good predictor of progression to AIDS or death, is not a good surrogate marker for the population of mildly symptomatic HIV-infected individuals.

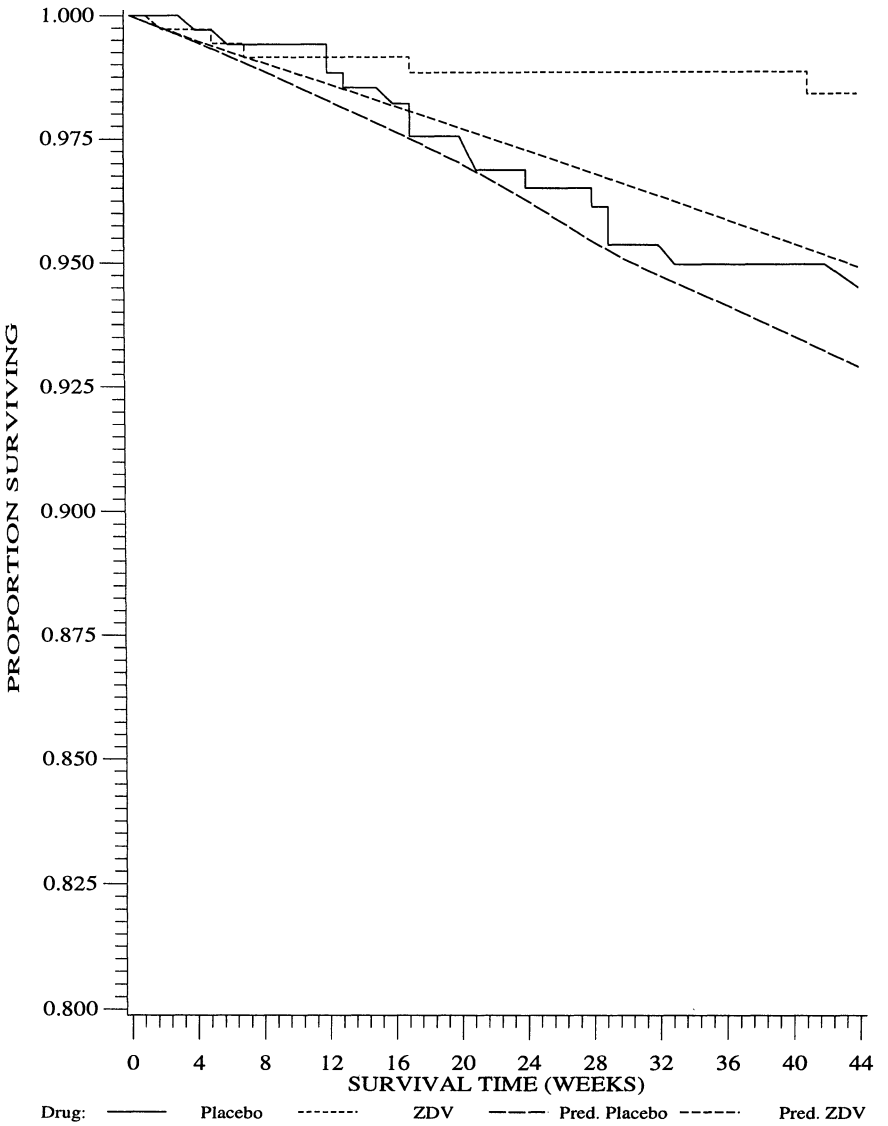

Figure 4. Observed survival and predicted improvement in survival resulting from ZDV-related increases in CD4-lymphocyte count for patients in ACTG 016.

The CD4 count has been the most promising and widely used candidate for a surrogate among the available immunological and virological markers. The consistency between the results of this work and the papers cited confirms that CD4 count cannot be successfully used as a surrogate for clinical outcome in HIV-infected individuals whether asymptomatic, mildly symptomatic, or with advanced disease. A combination of virological and immunological markers might be required to fulfill this role successfully. Further work is needed to evaluate batteries of markers. The ACTG large-scale comparative trials, with their wealth of information, can provide the data necessary for this evaluation.

6. Concluding Remarks

Our results indicate that using the naive approach of including the observed marker values measured periodically with error in the Cox model to evaluate a potential surrogate can lead to erroneous conclusions when testing either of the two empirical conditions set by Prentice (1989). Not only is the estimate of the association of the marker to survival biased toward zero, but the residual effect

that should have been captured by the marker can often be large enough to inflate the treatment effect to significance, even in the case of no missing data.

Our approach has the limitation of being based on distributional assumptions and a model for the marker time progression. For that reason, though better than the naive approach in discriminating between good and bad surrogate markers if the assumptions are true, we propose that our approach be used in combination with the naive approach. If neither approach identifies a marker as a good surrogate, then one can be more confident that the conclusion reached is the correct one. If they disagree, and our method identifies a marker as a good surrogate while the naive approach does not, caution should be exercised, and both the distributional assumptions and the model used for time progression should be checked. If the assumptions are verified, our approach is likely to have correctly identified the marker as a good surrogate.

Acknowledgements

This work was supported by National Cancer Institute grant CA-51962 and National Institute of Allergy and Infectious Diseases grants AI-31789 and AI-38855. The authors thank the referees and the associate editor for constructive suggestions.

Résumé

Dans la plupart des essais thérapeutiques, il existe une erreur associée à la mesure périodique des marqueurs. En présence d'erreur de mesure, l'introduction naive dans le modèle de Cox des valeurs observées du marqueur, afin d'évaluer son rôle sur le pronostic, peut conduire à des estimateurs biaisés et aboutir à des conclusions inexactes quand on évalue un marqueur intermédiaire potentiel. Nous proposons une approche en 2 étapes pour prendre en compte l'erreur de mesure et réduire le biais de l'estimation. Au cours de la première étape, un estimateur empirique Bayesien de la covariable dépendant du temps est calculé à chaque temps d'événement. Au cours de la deuxième étape, ces estimateurs sont introduits dans le modèle de Cox à risques proportionnels, afin d'estimer le coefficient de régression d'intérêt. Nous démontrons, par de multiples simulations, que cette méthodologie diminue le biais de l'estimateur de régression et identifie, plus souvent que l'approche naive, correctement les bons marqueurs intermédiaires. Un exemple est présenté, évaluant, dans un essai thérapeutique sur le SIDA, le nombre de CD4 comme marqueur intermédiaire de progression de la maladie.

References

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30,** 89–99.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models.* London: Chapman and Hall.

Choi, S., Lagakos, S. W., Schooley, R., and Volberding, P. (1993). CD4 lymphocytes are incomplete surrogate markers for clinical progression in persons with asymptomatic HIV infection taking Zidovudine. *Annals of Internal Medicine* **118,** 674–680.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B* **34,** 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

DeGruttola, V., Lange, N., and Dafni, U. (1991). Modeling the progression of HIV infection. *Journal of the American Statistical Association* **86,** 569–577.

Fischl, M. A., Richman, D. D., Hansen, N., et al. (1990). The safety and efficacy of Zidovudine (AZT) in the treatment of subjects with mildly symptomatic human immunodeficiency virus type 1 (HIV) infection. *Annals of Internal Medicine* **112,** 727–737.

Guttman, I. (1982). *Linear Models.* New York: Wiley.

Hughes, M. D. (1993). Regression dilution in the proportional hazards model. *Biometrics* **49,** 1056–1066.

Jones, R. H. and Ackerson, L. M. (1990). Serial correlation in unequally spaced longitudinal data. *Biometrika* **77,** 721–731.

Lagakos, S. W. and Hoth, D. F. (1992). Surrogate markers in AIDS: Where are we? Where are we going? *Annals of Internal Medicine* **116,** 599–601.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38,** 963–974.

Pepe, M. S., Self, G. S., and Prentice, R. L. (1989). Further results on covariate measurement errors in cohort studies with time response data. *Statistics in Medicine* **8,** 1167–1178.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69,** 331–342.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine* **8,** 431–440.

Prentice, R. L. and Self, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *The Annals of Statistics* **11,** 804–813.

Tsiatis, A. A., Dafni, U., DeGruttola, V., Propert, K. J., Strawderman, R. L., and Wulfsohn, M. (1992). The relationship of CD4 counts over time to survival in patients with AIDS: Is CD4 a good surrogate marker? In *AIDS Epidemiology: Methodological Issues*, N. Jewell, K. Dietz, and V. Farewell (eds), 256–274. Boston: Birkhauser.

Tsiatis, A. A., DeGruttola, V., and Wulfsohn, M. (1995). Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* **90,** 27–37.

Vittinghoff, E., Malani, H. M., and Jewell, N. P. (1994). Estimating patterns of CD4 lymphocyte decline using data from a prevalent cohort of HIV infected individuals. *Statistics in Medicine* **13,** 1101–1118.

Received May 1994; *revised March* 1997; *accepted December* 1997.

Appendix

At each time t and treatment k, an estimate for θ_k, Θ_k, σ_ϵ^2 is needed. The estimates are treatment specific. At time t, each individual belonging to $R_k(t)$ has a history of observed values for the marker $z_i^H(t)$. Only individuals on treatment k are included in the risk set $R_k(t)$. A regression line is fitted for each of the p_k surviving individuals on treatment k at time t, resulting in p_k least-squares estimates. We use as an estimate of θ_k, $\hat{\theta}_k$ the average of the p_k least-squares estimates:

$$\hat{\theta}_k^{(t)} = \frac{1}{p_k} \sum_{i=1}^{p_k} \left(W_i^{(t)'} W_i^{(t)} \right)^{-1} W_i^{(t)'} z_i^H(t), \tag{1}$$

where

$$W_i^{(t)} = \begin{bmatrix} 1, 1, \cdots, 1 \\ t_{i1}, t_{i2}, \cdots, t_{im} \end{bmatrix}',$$

$z_i^H(t) = (z_i(t_{i1}), z_i(t_{i2}), \ldots, z_i(t_{im}))'$, and $j = 1, \ldots, m$ indexes the times of observing the marker up to time t for individual i.

As an estimate of σ_ϵ^2, a weighted average of the p_k individual estimates of the residual sum of squares is used:

$$\hat{\sigma_{k\epsilon}^{2^{(t)}}} = \sum_{i=1}^{p_k} \frac{m_i - 2}{\sum_{i=1}^{p_k}(m_i - 2)} \hat{\sigma_{k\epsilon}^{2^{(i)}}} = \sum_{i=1}^{p_k} \frac{m_i - 2}{\sum_{i=1}^{p_k}(m_i - 2)} \left(\frac{RSS_i}{m_i - 2} \right) = \frac{\sum_{i=1}^{p_k} RSS_i}{\sum_{i=1}^{p_k}(m_i - 2)}. \tag{2}$$

To estimate the variance–covariance matrix of $(\alpha_{0ik}, \alpha_{1ik})$, $\Theta_k^{(t)}$ we make use of the conditional variance formula:

$$\operatorname{var}\begin{pmatrix} \hat{\alpha}_{0ik} \\ \hat{\alpha}_{1ik} \end{pmatrix} = \operatorname{var}\left[\mathrm{E}\left(\begin{pmatrix} \hat{\alpha}_{0ik} \\ \hat{\alpha}_{1ik} \end{pmatrix} \middle| \begin{pmatrix} \alpha_{0ik} \\ \alpha_{1ik} \end{pmatrix} \right) \right] + \mathrm{E}\left[\operatorname{var}\left(\begin{pmatrix} \hat{\alpha}_{0ik} \\ \hat{\alpha}_{1ik} \end{pmatrix} \middle| \begin{pmatrix} \alpha_{0ik} \\ \alpha_{1ik} \end{pmatrix} \right) \right]. \tag{3}$$

But

$$\begin{aligned} &\operatorname{var}\left[\mathrm{E}\left(\begin{pmatrix} \hat{\alpha}_{0ik} \\ \hat{\alpha}_{1ik} \end{pmatrix} \middle| \begin{pmatrix} \alpha_{0ik} \\ \alpha_{1ik} \end{pmatrix} \right) \right] \\ &\quad = \operatorname{var}\left[\mathrm{E}\left[\mathrm{E}\left(\begin{pmatrix} \hat{\alpha}_{0ik} \\ \hat{\alpha}_{1ik} \end{pmatrix} \middle| \begin{pmatrix} \alpha_{0ik} \\ \alpha_{1ik} \\ W_i \end{pmatrix} \right) \middle| \begin{pmatrix} \alpha_{0ik} \\ \alpha_{1ik} \end{pmatrix} \right] \right] \\ &\quad = \operatorname{var}\left[\mathrm{E}\left(\begin{pmatrix} \alpha_{0ik} \\ \alpha_{1ik} \end{pmatrix} \middle| \begin{pmatrix} \alpha_{0ik} \\ \alpha_{1ik} \end{pmatrix} \right) \right] = \operatorname{var}\begin{pmatrix} \alpha_{0ik} \\ \alpha_{1ik} \end{pmatrix} = \Theta_k. \end{aligned} \tag{4}$$

We estimate

$$\operatorname{var}\begin{pmatrix}\hat{\alpha}_{0ik}\\ \hat{\alpha}_{1ik}\end{pmatrix}$$

by the empirical variance of $\hat{\alpha}_{0ik}$, $\hat{\alpha}_{1ik}$:

$$\widehat{\operatorname{var}}\begin{pmatrix}\hat{\alpha}_{0ik}\\ \hat{\alpha}_{1ik}\end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{p_k}\left(\hat{\alpha}_{0ik}-\frac{\sum_{i=1}^{p_k}\hat{\alpha}_{0ik}}{p_k}\right)^2 & \sum_{i=1}^{p_k}\left(\hat{\alpha}_{0ik}-\frac{\sum_{i=1}^{p_k}\hat{\alpha}_{0ik}}{p_k}\right)\left(\hat{\alpha}_{1ik}-\frac{\sum_{i=1}^{p_k}\hat{\alpha}_{1ik}}{p_k}\right) \\ \sum_{i=1}^{p_k}\left(\hat{\alpha}_{0ik}-\frac{\sum_{i=1}^{p_k}\hat{\alpha}_{0ik}}{p_k}\right)\left(\hat{\alpha}_{1ik}-\frac{\sum_{i=1}^{p_k}\hat{\alpha}_{1ik}}{p_k}\right) & \sum_{i=1}^{p_k}\left(\hat{\alpha}_{1ik}-\frac{\sum_{i=1}^{p_k}\hat{\alpha}_{1ik}}{p_k}\right)^2 \end{pmatrix} \tag{5}$$

and,

$$\mathrm{E}\left[\operatorname{var}\left(\begin{pmatrix}\hat{\alpha}_{0ik}\\ \hat{\alpha}_{1ik}\end{pmatrix}\middle|\begin{pmatrix}\alpha_{0ik}\\ \alpha_{1ik}\end{pmatrix}\right)\right]$$

by the average of the p_k least-squares estimates, because for individual i:

$$\begin{aligned}
&\widehat{\operatorname{var}}\left(\begin{pmatrix}\hat{\alpha}_{0ik}\\ \hat{\alpha}_{1ik}\end{pmatrix}\middle|\begin{pmatrix}\alpha_{0ik}\\ \alpha_{1ik}\end{pmatrix}\right)\\
&\quad=\mathrm{E}\left[\operatorname{var}\left(\begin{pmatrix}\hat{\alpha}_{0ik}\\ \hat{\alpha}_{1ik}\end{pmatrix}\middle|\begin{pmatrix}\alpha_{0ik}\\ \alpha_{1ik}\\ W_i\end{pmatrix}\right)\middle|\begin{pmatrix}\alpha_{0ik}\\ \alpha_{1ik}\end{pmatrix}\right]\\
&\quad+\operatorname{var}\left[\mathrm{E}\left(\begin{pmatrix}\hat{\alpha}_{0ik}\\ \hat{\alpha}_{1ik}\end{pmatrix}\middle|\begin{pmatrix}\alpha_{0ik}\\ \alpha_{1ik}\\ W_i\end{pmatrix}\right)\middle|\begin{pmatrix}\alpha_{0ik}\\ \alpha_{1ik}\end{pmatrix}\right]\\
&\quad=\mathrm{E}\left[\sigma_\epsilon^2\left(W_i^{(t)'}W_i^{(t)}\right)^{-1}\middle|\begin{pmatrix}\alpha_{0i}\\ \alpha_{1i}\end{pmatrix}\right]+0=\sigma_\epsilon^2\left(W_i^{(t)'}W_i^{(t)}\right)^{-1}.
\end{aligned}$$

Thus,

$$\hat{\mathrm{E}}\left[\operatorname{var}\left(\begin{pmatrix}\hat{\alpha}_{0i}\\ \hat{\alpha}_{1i}\end{pmatrix}\middle|\begin{pmatrix}\alpha_{0i}\\ \alpha_{1i}\end{pmatrix}\right)\right]=\frac{1}{p_k}\sum_{i=1}^{p_k}\hat{\sigma_k}_\epsilon^2\left(W_i^{(t)'}W_i^{(t)}\right)^{-1}=\frac{1}{p_k}\sum_{i=1}^{p_k}\frac{RSS_i}{m_i-2}\left(W_i^{(t)'}W_i^{(t)}\right)^{-1}. \tag{6}$$

Substituting the estimates (5) and (6) in (3), we get $\hat{\Theta}_k^{(t)}$, the estimate for the variance–covariance matrix of $(\alpha_{0ik}, \alpha_{1ik})$.