IDENTIFICATION AND ASSESSMENT OF LONGITUDINAL BIOMARKERS USING FRAILTY MODELS IN SURVIVAL ANALYSIS

by

Feng-shou Ko

MS, University of Pittsburgh, 2004

MS, National Yang-Ming University, 1997

BS, National Yang-Ming University, 1995

Submitted to the Graduate Faculty of

the Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Feng-shou Ko

It was defended on

July 24, 2006

and approved by

Dissertation Advisor:

Stewart J. Anderson, Ph.D, Associate Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Sati Mazumdar, Ph.D, Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Joseph P. Costantino, Dr.PH, Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Chyongchiou J. Lin, Ph.D, Associate Professor, Department of Radiation Oncology, School of Medicine, University of Pittsburgh

IDENTIFICATION AND ASSESSMENT OF LONGITUDINAL BIOMARKERS USING FRAILTY MODELS IN SURVIVAL ANALYSIS

Feng-shou Ko, PhD

University of Pittsburgh, 2006

A biomarker is a measurement which can be used as a predictor or sometimes even a surrogate for a biological endpoint that directly measures a patient's disease or survival status. Biomarkers are often measured over time and so are referred to as longitudinal biomarkers. Biomarkers are of public health interest because they can provide early detection of life threatening or fatal diseases.

It is important in public health to be able to identify biomarkers to predict survival for patients because it can reduce the time and cost necessary to resolve the study question or used to identify subsets of patients who would be appropriate candidates for the administration of a targeted therapy. In this dissertation, we introduce a method employing a frailty model to identify longitudinal biomarkers or surrogates for a time to event outcome. Our method is an extension of earlier work by Wulfson, Tsiatis, and Song where it was assumed that the event times have the same baseline hazard. In our method, we allow random effects to be present in both the longitudinal biomarker and underlying survival function. The random effect in the biomarker is introduced via an explicit term while the random effect in the underlying survival function is introduced by the inclusion of frailty parameters into the model. We use simulations to explore how the number of individuals, the number of time points per individual and the functional form of the random effects from the longitudinal biomarkers influence the power to detect the association of the longitudinal biomarker and

the survival time. We also explore effect of missingness on how a biomarker predicts a time to event outcome. We conclude that for a given sample size, the biomarker effectiveness for relatively small numbers of subjects and large numbers of observed time points is better than for relatively large numbers of subjects and small numbers of observed time points. We also conclude that when the missing data mechanism is missing at random (MAR), our method works reasonably well. However, when the missing data mechanism is non-ignorable, our method doesn't perform well in determining whether or not potential biomarkers are good predictors of a time to event outcome. Finally, we apply our method to liver cirrhosis data and conclude that prothrombin is a good predictor of time to liver cirrhosis and thus, can be used as a potential surrogate for liver failure.

Key Words: surrogate; biomarker; multivariate survival; frailty model.

TABLE OF CONTENTS

LIST OF TABLES

1.0 INTRODUCTION

A biological characteristic that is a surrogate for an outcome of interest is referred to as a biomarker. In a clinical trial, it is often useful to be able to identify a biomarker as a surrogate for the outcome of interest because the use of the surrogate can reduce the time and cost necessary to resolve the study question or used to identify subsets of patients who would be appropriate candidates for the administration of a targeted therapy. Longitudinal data analysis and survival analysis have been used broadly as individual methods to analyze biological and medical studies. In recent years, some investigators have employed these two statistical methodologies in a combined approach to analysis. Joint modeling methods are the potential to exploit the longitudinal biomarker as a surrogate for the subsequent survival. Individual-level surrogacy for a survival endpoint will be focused in this study. A score test for association between survival time and biomarker values is developed.

Traditionally, a longitudinal biomarker is used for monitoring survival based on the assumption that the baseline hazard for each observation is homogeneous. However, many studies show that the assumption that the survival time for each observation is homogeneous is not adequate. For example, the survival analysis of twins study, the occurrence of the events such as the death is not based on the homogeneous baseline hazard. Other examples include the study of the diabetic retinopathy and the recurrent event study, the occurrence of the event is also not based on the homogeneous baseline hazard. Hence, it is necessary to develop a more adequate method to deal with the studies about the longitudinal biomarker as the surrogate, which is used for monitoring the survival situation for the patients when the association of the observation times is dependent.

In recent years, frailty models are used to deal with a heterogenous hazard in survival analysis. Frailty models are basically random effects models for the survival data, where the random effects are specified by means of the hazard function. Gamma, Weibull, and lognormal distribution are usually assumed as the distribution of frailty models.

The efficacy of individual-level biomarkers is also considered. From one point of view, an effective surrogate is one for which the conditional residual lifetime distribution accounting for biomarkers information at an interim time is more strongly concentrated around the actual, but as-yet unobserved, survival time than is the marginal residual lifetime distribution ignoring the biomarker data. When a biomarker is considered to be a surrogate, a statistical method is necessary to determine the effectiveness of the surrogate. An effective surrogate is very useful for the investigator to predict the patients' survival status. In survival analysis, predictive accuracy for the individual patients should be distinguished from the accuracy of survival function estimates. The appropriate method to determine an effective surrogate is very important for the multivariate survival analysis.

In this dissertation, we will try to address the following problems:

1. How to determine a longitudinal biomarker as a surrogate for survival with heterogeneous hazard
2. How to deal with measurement errors for the covariates in the Cox model
3. Finding a joint model for the combined analysis of survival with heterogeneous hazard and longitudinal data
4. Effectiveness of biomarkers as surrogate endpoints

Accordingly, the specific goals of this dissertation are to:

1. Extend the method proposed by Henderson, Diggle, and Dobson (2002) for survival with heterogeneous hazard data will be developed.
2. Develop the joint likelihood function which combines the likelihood functions of the longitudinal biomarkers and survival with heterogeneous hazard distribution.
3. Develop a survival with heterogeneous hazard model based on a frailty model.
4. Incorporate frailty into the semi-parametric hazard model (Cox model); and

5. Develop the methodology to determine the effectiveness of potential longitudinal biomarkers for survival with heterogeneous hazard data.

2.0 REVIEW OF KEY LITERATURE

2.1 LONGITUDINAL BIOMARKER AS A SURROGATE

The idea of using one endpoint as a surrogate for a later occurring endpoint has been explored extensively over the last 25 years. Paterson, et al. (1985) [1] discussed the association between the response to treatment and survival of patients who have metastatic breast cancer. In their study, the response to treatment was classified into complete response, partial response, stable disease, and progressive disease. They found that survival time among the patients with different responses to treatment were not significantly different even when controlling for menopausal status and treatment method except patients who were classified into progressive disease. Patients who were classified as having progressive disease had a shortened survival time. They suggested that the assessment of a treatments worth should be based as much on the patients subjective feeling of well-being as on the magnitude of the tumor response.

Gail [2] published a paper evaluating serial cancer markers in patients at risk of recurrent disease. He showed that high levels of carcinoembryonic antigen (CEA) are associated with increased risk of death in patients with resected colorectal cancer. He also defined time dependent functions, Z (t), which summarize the marker history up to time t. These functions were tested to determine whether the marker was related to risk of death (or recurrence). A parameter, γ, denoted the value of a marker at time t and the marker history for an individual was defined as M = $\{\gamma(t) : 0 \leq \tau \leq t\}$. He specified the following definitions of Z (t):

Z_1 (t) $= \gamma(t)$;

Z_2 (t) $= \gamma(t - \omega)$;

Z_3 (t) $= 1$ if sup $\gamma(t) \geq \eta, 0 \leq \tau \leq t$, 0 otherwise; and

Z_4 (t) $= \{\gamma(t) - \gamma(t - \omega)\}/\Delta$.

Z_1 (t) is the simplest function and may be used to study the question, "Are those with elevated marker values at time t at higher risk at a given time than those without elevated marker values?" Z_2 (t) may be used to test whether those with elevated marker values at time $t - \omega$ at higher risk at time t. Z_3 (t) could be used to assess if any previous elevation of the marker increases risk, and Z_4 (t) could be used to see if a high rate of increase in the marker has grave prognostic significance. Gail commented about some generic features of the serial marker data problem:

(1) The risk of death may be influenced by other prognostic factors which could obscure the effect of the serial marker;

(2) There is often insufficient information to justify a particular parametric model for an analysis;

(3) The time to death data are variably censored on the right;

(4) The marker value $\gamma(t)$ is only measured at a finite number of points; and

(5) Occasionally one has no idea what value to assign $\gamma(t)$ or Z (t) because no proximate values are available.

Gail provided some methods to solve these problems. The first three problems arise whenever one attempts a covariate analysis on survival data, and the semiparametric approach of Cox is well adapted to these problems. Other prognostic factors can be adjusted by stratification and allow a separate nuisance hazard function for each stratum. Problem (4) requires the data analyst to define an interpolation convention to assign values $\gamma(t)$ for t intermediate times between observations. The partial likelihood ratio method of Cox is particularly useful for problem (5), because it allows patients to contribute to "risk sets" when, and only when, a valid marker measurement is available. "Proximate" in (5) means we're interested in the neighborhood area around the interest of time point t or t - ω. That

is, the interval of $(t-\Delta t, t+\Delta t)$ or $(t-\omega-\Delta t, t-\omega+\Delta t)$. A marker measurement, $\gamma(t)$, is said to be censored if no proximate measurements are available; otherwise, $\gamma(t)$ is said to be a valid marker measurement. Likewise, the experiment yields a valid marker measurement or censored functional measurement according as sufficient proximate observations are or are not available to determine Z (t). For the same set of marker measurements, some functions, such as the slope Z_4(t), may be censored and others, such as Z_1(t) may be valid.

Tsiatis, DeGruttola, and Wulfsohn tried to find a good surrogate marker to evaluate new treatments in acquired immune deficiency syndrome (AIDS) clinical trials [3]. The Cox proportional hazards regression model was used to study the relationship between CD4 counts as a time-dependent covariate and survival. The authors indicated that a good surrogate biomarker should have the following properties: 1) it should be related to prognosis; 2) the distribution of the values for the biomarker should be different for individuals receiving an effective treatment versus those receiving a placebo; and 3) the beneficial effects of a good treatment should be mediated through its effect on the marker. In other words, patients with the same value of a biomarker should have the same prognosis whether they are receiving a treatment or a placebo. In such a case, the better prognosis associated with a good treatment could be explained by the change in the value of the marker for that treatment. The authors also commented that the standard methods for estimating the parameters in the Cox model by maximizing the partial likelihood are not appropriate because the CD4 counts are measured only periodically and with substantial measurement error because of biological variation. They proposed a two-stage method approach to estimate the parameters. In the first stage, the longitudinal CD4 count data are modeled using a repeated measures random components model. In the second stage, methods for estimating the parameters in a Cox model when the data are assumed to be of this form are derived. Tsiatis, DeGruttola, and Wulfsohn also used the new methods to deal with the questions about the missing data. They analyzed the CD4 data from a randomized clinical trial of AIDS patients where half of the patients were randomized to receive Zidovudine (ZDV) and the other half of them were randomized to receive a placebo. The results of the study showed that the CD4 counts

might not serve as a useful surrogate biomarker for assessing treatments for the population of patients.

Weiss, Bunce, and Hokanson commented that the comparison of survival distributions between responding and non-responding patients can be difficult in interpretation and methodology [4]. The statistical test only shows the association between response and survival but it does not mean that the relationship of the cause and effect exists since the assignment of patients into groups is not random. Because this association might have no relevance to the efficacy of treatment, it is difficult to interpret the efficacy of treatment by the comparing the survival distributions of responding and non-responding patients. Besides, variability in the definition of a non-responder and the handling of early deaths could cause different conclusions concerning survival.

Wittes, Lakatos, and Probstfield discussed surrogate endpoints for cardiovascular disease in clinical trials [5]. They defined a surrogate endpoint as an endpoint measured as an alternative to some other "true endpoint". A surrogate is especially useful if it is easily measured and highly correlated with the true endpoint. Often, the true endpoint is one with clinical importance to the patient, for example, mortality or a major clinical outcome, while a surrogate is one biologically closer to the process of disease, for example, cardiac ejection fraction. Use of the surrogate can often lead to dramatic reductions in sample size and much shorter studies than use of the true endpoint. Several problems common in trials with surrogate endpoints are discussed in the paper. Most important is the effect of missing data, especially in the face of informative censoring. Wittes, Lakatos, and Probstfield suggested three possible methods for dealing with missing endpoints in clinical trial. First, analyze the data available and ignore the fact that some observations are missing. This is the most common approach and has large potential bias. Second, use a formal statistical method to attempt to reduce the bias caused by informative censoring. The simplest approach is to assign a score to the missing value. More complicated methods are under investigation, but as mentioned above, have little practical use if a large proportion of data is missing. Third, use an informal rule to penalize a study with missing data or as part of a sensitivity analysis.

They also commented on the heterogeneity of variance as another problem common to many surrogate endpoints and commented that the required sample size is often too small to detect infrequent but major adverse effects of therapy.

A major obstacle in the study of the etiology of chronic disease and the development of effective prevention is the long latent period between the initiation of the disease and its diagnosis. Prospective studies relating possible risk factors to disease or investigating the effects of an intervention on disease incidence therefore require extended periods of follow-up. Such studies are costly. Freedman, Graubard, and Schatzkin [6] defined intermediate endpoints (IE) that are biological markers or events that may be assessed or observed prior to the clinical appearance of the disease, and that bear some relationship to the development of that disease. In the study of chronic disease, the use of IE can shorten the duration of follow-up time needed to assess the efficacy of an intervention or the association of a risk factor with outcome. They listed four points about how intermediate points may be studied and validated. First, intermediate endpoints should usually be validated within prospective studies, either observational cohort studies or experimental intervention trials. Second, in a cohort study we need to examine the exposure–IE–disease relationship; in an intervention study the intervention–IE–disease relationship should be examined. Third, intermediate endpoints for a disease can only be validated in reference to a given exposure (or intervention). Once validated for that exposure the IE may be considered valid for other exposures that affect the disease through the same pathway. Fourth, the criterion for validation is that the exposure (or intervention) effect on disease, adjusted for the intermediate endpoint, is equal to zero. Freedman, Graubard, and Schatzkin also analyzed the data from the lipid research clinics coronary primary prevention trial to examine whether serum cholesterol level is an intermediate endpoint for coronary heart disease (CHD) by investigating the effect of the cholesterol lowering drug cholestyramine on CHD incidence adjusted for serum cholesterol levels. They found serum cholesterol level is not a good intermediate endpoint for coronary heart disease (CHD).

In 1992, Pepe [7] discussed inference using surrogate outcome data and a validation sample. In her study, most subjects only had covariates and surrogate response data. Few subjects had complete data including true outcome response. Parametric and semi-parametric methods were used to estimate such data.

Buyse, et al. [8] discussed validating surrogate endpoints via meta-analyses of randomized experiments and commented on the definition of a surrogate as presented by Prentice [9]. Prentice had defined a surrogate endpoint as response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint. If T and S are random variables that denote the true and surrogate endpoints, respectively, and Z is an indicator variable for treatment, then, using the notation of Buyse et al., Prentice's definition can be written as follows: f (S| Z) = f (S) iff f (S| T) = f (S), where f (S) denotes the probability distribution of the surrogate endpoint, f (S| T) denotes the probability distribution of the surrogate endpoint conditional on the value of the true endpoint, and f (S| Z) denotes the probability distribution of S conditional on the value of Z, an indicator variable for treatment. As such, the definition is of limited value since a direct verification that a triplet (T, S, Z) fulfills the definition would require a large number of experiments to be available with information on the triplet. Operational criteria are therefore needed to check if Prentices definition is fulfilled. Buyse et al., commented that four operational criteria have been proposed to check if the triplet (T, S, Z) fulfills the definition. The first two criteria are: 1) f (S|Z) $\neq$ f (S) and; 2) f (T|Z) $\neq$ f (T). Both of these criteria are consistent with Prentice's definition. In practice, due to lack of power, the validation of these criteria requires Z to have an effect on both T and S. It has been pointed out that requiring Z to have a statistically significant effect on T may be excessively stringent, because in that case, from the limited perspective of significance testing, there would no longer be a need to establish the surrogacy of S. The two other criteria are: 3) f (T|S) $\neq$ f (T) and 4) f (T|S, Z) = f (T|S). It can be proven that criterion #3 is sufficient for Prentices definition in all cases, and criterion #4 is sufficient for binary endpoints but not in general. These four operational

criteria are informative and will tend to be fulfilled for valid surrogate endpoints, but they should not be regarded as strict criteria. Criterion #4, f (T|S, Z) = f (T|S) captures the essential notion of surrogacy by requiring that the treatment is irrelevant for predicting the true outcome, given the surrogate. Buyse et al. commented that f (T|S, Z) = f (T|S) raises a conceptual difficulty in that it requires the statistical test for treatment effect on the true endpoint to be non-significant after adjustment for the surrogate. The non-significance of this test does not prove that the effect of treatment upon the true endpoint is fully captured by the surrogate. However, it is proposed to calculate the proportion of the treatment effect explained by surrogate. A good surrogate is one for which this proportion explained (PE) is close to unity based on this concept (f (T|S, Z) = f (T|S) would require that PE = 1). Then they argued this concept and proposed an alternative. They used two related quantities to replace PE. One is the relative effect (RE) that is the ratio of the effects of treatment upon the final and the surrogate endpoint. The other is the treatment-adjusted association between the surrogate and the true endpoint, ρ_z.

Buyse and Molengerghs commented criteria for the validation of surrogate endpoints in randomized experiments [10]. They focused on the cases where the surrogate and the final endpoints were both binary and normally distributed. Letting T and S be random variables that denote the true and surrogate endpoint, respectively, and Z be an indicator variable for treatment, Prentice's criteria are fulfilled if Z has a significant effect on T and on S, if S has a significant effect on T, and if Z has no effect on T given S. Freedman relaxed the latter criterion by estimating PE, the proportion of the effect of Z on T that is explained by S, and by requiring that the lower confidence limit of PE be larger than some proportion, say 0.5 or 0.75. This condition can only be verified if the treatment has a massively significant effect on the true endpoint, a rare situation. They argued that two other quantities must be considered in the validation of a surrogate endpoint: RE, the effect of Z on T relative to that of Z on S, and γ_z, the association between S and T after adjustment for Z. A surrogate is said to be perfect at the individual level when there is perfect association between the surrogate and the final endpoint after adjustment for treatment. A surrogate is said to be

perfect at the population level if RE is 1. A perfect surrogate fulfills both conditions, in which case S and T are identical up to a deterministic transformation. Fieller's theorem was used for the estimation of PE, RE, and their respective confidence intervals. Logistic regression models and the global odds ratio models were used for binary endpoints. Linear models were employed for continuous endpoints. In order to be of practical values, the validation of surrogate endpoints was shown to require large numbers of observations.

In 2000, Gail, et al., discussed the strengths and weaknesses of the meta-analytic approach to estimate the effect of a new treatment on a true clinical outcome measure, T, from the effect of treatment on a surrogate response, S [11]. The meta-analytic approach uses data from a series of previous studies of interventions similar to the new treatment. The data are used to estimate relationships between summary measures of treatment effects on T and S that can be used to infer the magnitude of the effect of the new treatment on T from its effects on S. The class of models is extended to cover a broad range of applications in which the parameters define features of the marginal distribution of (T, S). A bootstrap procedure is also presented to allow for the variability in estimating the distribution that governs the between-study variation. Gail, et al. noted that ignoring this variability can lead to confidence intervals that are much too narrow. They also noted that, compared to direct measurement on T, the meta-analytic approach has limitations including the likely serious loss of precision and difficulties in defining the class of previous studies to be used to predict the effects on T for a new intervention.

Bruzzi commented on phase II studies that used tumor response to chemotherapy as the primary endpoint to evaluate the anti-tumor activity of new drugs [12]. He concluded that tumor response is indeed a valid surrogate endpoint of survival in colorectal cancer, and that there is strong indirect evidence supporting a similar role of tumor response in breast cancer. The author commented that this biomarker may be a good candidate for use as a surrogate in the trial of metastatic breast cancer to aid in decision for testing patients.

2.2 MEASUREMENT ERRORS FOR THE COVARIATES IN THE COX MODEL

Hu, Tsiatis, Davidian commented that estimating the parameters in the Cox model when covariate variable are measured with error [13]. The Cox proportion hazards model is commonly used to model survival data as a function of covariates. Because of the measuring mechanism or the nature of the environment, covariates are often measured with error and are not directly observable. A naive approach is to use the observed values of the covariates in the Cox model, which usually produces biased estimates of the true association of interest. An alternative strategy is to take into account the error in measurement, which may be carried out for the Cox model in a number of ways. They examined several such approaches and compare and contrast them through several simulation studies. They introduced a likelihood-based approach, which they referred to as the semiparametric method, and showed that this method was an appealing alternative. The methods were applied to analyze the relationship between survival and CD4 count in patients with AIDS.

Tsiatis and Davidian [14] discussed a semiparametric estimation for the proportional hazards model with longitudinal covariates measured with error. They commented that a common objective in longitudinal studies is to characterise the relationship between a failure time process and time-dependent covariates are generally available as longitudinal data collected periodically during the course of the study. They assumed that these data follow a linear mixed effects model with normal measurement error and that the hazard of failure depends both on the underlying random effects describing the covariate process and other time-independent covariates through a proportional hazards relationship. A routine assumption is that the random effects are normally distributed; however, this need not hold in practice. Within this framework, they developed a simple method for estimating the proportional hazards model parameters that required no assumption on the distribution of the random effects. Large-sample properties were discussed, and finite-sample performance is assessed and compared to competing methods via simulation.

Liu, Mazumdar, Stone, Dew, Houck, Reynolds [15] commented Accounting for covariate measurement error in a Cox model analysis of recurrence of depression. When a covariate measured with error is used as a predictor in a survival analysis using the Cox model, the parameter estimate is usually biased. In clinical research, covariates measured without error such as treatment procedure or sex are often used in conjunction with a covariate measured with error. In a randomized clinical trial of two types of treatment, we account for the measurement error in the covariate, log-transformed total rapid eye movement (REM) activity counts, in a Cox model analysis of the time to recurrence of major depression in an elderly population. Regression calibration and two variants of a likelihood-based approach are used to account for measurement error. The likelihood-based approach is extended to account for the correlation between replicate measures of the covariate. Using the replicate date decreases the standard error of the parameter estimate for correlation between replicates can affect results in a Cox model analysis and should be accounted for. In the depression data, these methods render comparable results that have less bias than the results when measurement error is ignored.

2.3 JOINT MODELS FOR SURVIVAL AND LONGITUDINAL DATA ANALYSIS

Wulfsohn and Tsiatis discussed a joint model for survival and longitudinal data measured with error [16]. They commented that the relationship between a longitudinal covariate and a failure time process can be assessed using the Cox proportional hazards regression model. They considered the problem of estimating the parameters in the Cox model when the longitudinal covariate is measured infrequently and with measurement error. They assumed a repeated measures random effects model for the covariate process. Estimates of the parameters were obtained by maximizing the joint likelihood for the covariate process and the failure time process. This approach used the available information optimally because

they use both the covariate and survival data simultaneously. Parameters were estimated using the expectation-maximization algorithm. They argued that such a method is superior to naive methods where one maximizes the partial likelihood of the Cox model using the observed covariate values. It also improves on two-stage methods where, in the first stage, empirical Bayes estimates of the covariate process were computed and then used as time-dependent covariates in a second stage to find the parameters in the Cox model that maximize the partial likelihood.

Henderson, Diggle, and Dobson discussed the joint modeling of longitudinal measurements and event time data [17]. A class of models is formulated for the joint behavior of a sequence of longitudinal measurements and an associated sequence of event times, including single-event survival data. Special cases of the model class are discussed in detail and an estimation procedure which allows the two components to be linked through a latent stochastic process is described.

Huang and Louis commented nonparametric estimation of the joint distribution of survival time and mark variables [18]. In many applications, variables of interest were marks of the endpoint which were not observed when the survival time was censored. They focused on nonparametric estimation of the joint distribution and summaries of survival time and mark variables. They established a representation of the joint distribution function through the cumulative mark-specific hazard function, which was analogous to the product integral representation of univariate survival function. They identified a basic structure common to various applications, proposed nonparametric estimators and showed that they examined the likelihood. They formulate the problem in the marked point process framework and study both finite and large-sample properties of the estimators. We showed that the joint distribution function estimator was nearly unbiased, uniformly strongly consistent and asymptotically normal. They also derived asymptotic variances for the estimators and propose sample-based variance estimates. Numerical studies demonstrated that both the estimators and their variance estimates performed well for practical sample sizes. They outline an application strategy.

Lin, Turnbull, McCulloch, and Slate commented latent class models for joint analysis of longitudinal biomarker and event process data [19]. They commented that latent class models that incorporated both a longitudinal biomarker process and an event process offered a way to handle additional heterogeneity, to uncover distinct subpopulations, to incorporate correlated nonnormally distributed outcomes, and to classify individuals into risk classes. Their latent class joint model can aid the prediction of outcome variable probability given the longitudinal biomarker information available on an individual up to any date. The proposed model easily accommodated highly unbalanced longitudinal data and recurrent events. There were two levels of structure in the latent class joint model. First, the uncertainty of latent class membership was specified through a multinomial logistic model. Second, the class-specific marker trajectory and event process were specified parametrically and semiparametrically, under the assumption of conditional independence given the latent class membership. They used a likelihood approach to obtain parameter estimates via the EM algorithm.

2.4 EFFECTIVENESS OF A BIOMARKER AS A SURROGATE

Van der Laan, Hubbard, and Robin discussed locally efficient estimation of a multivariate survival function in longitudinal studies [20]. They considered estimation of the joint distribution of multivariate survival times T = $(T_1, \ldots, T_k)$, which were subject to right censoring by a common censoring variables C. Two estimators were proposed: an initial inverse-probability-of-censoring weighted (IPCW) estimator, and a 1-step estimator. Both estimators incorporated information on available time-independent and time-dependent prognostic factor (covariate) data. The IPCW estimator was consistent and asymptotically normal (CAN) under coarsening at random (CAR) and a correct specification of a model for the hazard of censoring given the past covariate and failure data. The 1-step estimator was a locally efficient doubly robust estimator. That is, (i) it was CAN under the assumption of CAR and either (but not necessarily both) correct specification of a model for the hazard

of censoring given the past or correct specification of a model for the conditional distribution of T given past failure and covariate information, and (ii) it was efficient when both these models are correctly specified. The proposed methodology did not required that the time variables $T_1, \ldots, T_k$ should be ordered, although their methods covered this important special case. In particular, their estimators can be used to estimate the gap time distributions associated with an ordered series of events. The proposed methodology improved over currently available approached in a number of ways. Specially, when censoring and failure were dependent because the hazard of censoring depended on both past failure and covariate history, our one-step estimator is the only estimator with the double robustness property. When censoring can be assumed to be independent of the failure and covariate process, our locally efficient one-step estimator did not require smoothing and so will perform well in moderate size samples even if k is large; further unlike all previous estimators, their estimator exploited the information available in past covariate as well as failure history and so will be efficient (nearly efficient) even when the components of T were highly dependent, whenever the specified model for the conditional distribution of T given past failure and covariate information was correct (nearly correct).

In 2003, Dobson and Henderson commented diagnostic for joint longitudinal and dropout time modeling [21]. There were three aims to their paper. The first was to propose an exploratory method designed to assess whether there was any association between responses and dropout time before any sophisticated and computationally intensive joint modeling was carried out. They argued that if no systematic differences between subjects who did or did not drop out can be found in the observed data, then joint modeling is unlikely to be worthwhile. This idea can be extended to investigate differences between subjects with different reasons for dropout. Often the reason for dropout from a longitudinal trial was not stated, but sometimes a reason was given and can be classified as either potentially informative or otherwise. For instance, being too ill to continue was clearly informative, whereas leaving the study region may not be. They assumed that there were two categories of withdrawal, one which might be related to the unobserved response of interest (potentially

informative dropout), and one which was known or assumed to be independent of the unobserved response (assumed noninformative dropout). With minor modifications, the methods can be adapted either to unclassified dropout reasons or to situations where there were more than two dropout categories of interest. The second aim was to suggest conditional residual analysis methods for longitudinal data with dropout. They showed that residuals between observed and expected responses after fitting a joint model can be markedly affected by knowledge of the dropout time and type, which therefore should properly be taken into account in an assessment of model adequacy. The final aim was to advocate and illustrate ideas of case influence for joint modeling. Full case deletion was unrealistic in practice, because of the computing time required, and some form of approximation was essential. They presented a variety of informal graphical procedures for diagnostic assessment of joint models for longitudinal and dropout time data. A random effects approach for Gaussian response and proportional hazards dropout time was assumed. They considered preliminary assessment of dropout classification categories based on residuals following a standard longitudinal data analysis with no allowance for informative dropout. Residual properties conditional upon dropout information were discussed and case influence was considered. The proposed methods do not require computationally intensive methods over and above those used to fit the proposed model.

3.0 METHODS

3.1 MODEL AND NOTATION

The method of Henderson, Diggle, and Dobson [22] was used to analyze follow-up data dealing with the longitudinal measurement of a time-varying biomarker. They let $\mathbf{Y}_i(t)$ represent the underlying biomarker vector of the i^{th} individual at time t, so that the equation can be written

$$\mathbf{Y}_i(t) = \mathbf{x}'_{i1}(t)\boldsymbol{\beta}_1 + \mathbf{W}_i(t) + \mathbf{e}_i(t).$$

where $\mathbf{x}_{i1}(t)$ is a $p_1 \times 1$ vector of explanatory variables, and $\mathbf{W}_i(t)$ and $\mathbf{e}_i(t)$ are zero-mean random processes. Since the biomarkers are sampled at discrete time points and assuming that the $\mathbf{Y}_i$ are from a normal distribution, a discrete version of the model can be re–written as

$$\mathbf{Y}_{ij} = \mathbf{x}'_{ij1}\boldsymbol{\beta}_1 + \mathbf{W}_{ij} + \mathbf{e}_{ij}, \quad i = 1, 2, \ldots, m; \; j = 1, 2, \ldots, n_i \tag{3.1}$$

where $\mathbf{W}_{ij}$ is the value of a zero-mean Gaussian random effect for the i^{th} individual at time j and $\mathbf{e}_{ij}$ is a zero-mean Gaussian measurement error. The j's in the discrete version of the model refer to the last biomarkers observed at or before time t. The errors, $\mathbf{e}_{ij}$, are assumed here to be mutually independent and the within individual correlation in $\mathbf{Y}_{ij}$ arises through serial correlation in the random effect, $\mathbf{W}_{ij}$.

It is assumed that survival time is associated with the longitudinal response through the effect of the latent process, $\mathbf{W}_{ij}$, but is otherwise conditionally independent. A semiparametric proportional hazards model is also assumed and the density function associated with the hazards model is presented as follows:

$$f(t_{ij}) = S(t_{ij})\lambda_0(t_{ij}) \exp\{\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma \mathbf{W}_{ij}\} \tag{3.2}$$

where $S(t_{ij})$ is a predictable survival function, $\lambda_0(t)$ is an unspecified baseline hazard, and $\mathbf{x}_{ij2}$ is a $p_2 \times 1$ vector of explanatory variables. The generic notation T, $\mathbf{Y}$ and $\mathbf{W}$ are used for survival time, the longitudinal response and the latent process, respectively.

3.2 A SCORE TEST FOR ASSOCIATION BETWEEN LONGITUDINAL BIOMARKER VALUES AND SURVIVAL TIME FUNCTION

The score test used for testing the parameters of equation (3.2) and is based on separate analyses of $\mathbf{Y}$ and T under the null hypothesis, $H_0 : \gamma = 0$ and it is assumed that $\mathbf{Y}$ is multivariate Gaussian and T follows a proportional hazards model. Let the combined vector of unknown parameters be $(\boldsymbol{\theta}, \gamma, \boldsymbol{\beta}_2, A_0)$, where $\boldsymbol{\theta}$ contains all parameters of the distribution of $\mathbf{Y}$ and let the maximum follow-up time be t. $A_0(t)$ denotes the cumulative baseline hazard. In practice, the usual maximum partial likelihood estimator $\hat{\boldsymbol{\beta}}_2$ replaces the unknown $\boldsymbol{\beta}_2$. A_0 (t) is replaced by the non-parametric maximum likelihood estimator (under H_0) as follows:

$$\hat{A}_0(t) = \int_0^t \frac{J(u)}{\sum_{i=1}^m S_i(u) e^{\mathbf{x}'_{ij2}\beta_2}} dN(u)$$

where $N(u) = \sum N_i(u)$ and $J(u) = I(\sum S_i(u) > 0)$. If $\mathbf{W}$ is known, the conditional likelihood of the survival data can be written as follows:

$$L_\gamma = (\prod_j \prod_i (e^{\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma \mathbf{W}_{ij}})^{\Delta N_i(t)}) \exp\left\{ - \int_0^\tau S_\gamma^{(0)}(t, \mathbf{W}, \boldsymbol{\beta}_2) dA_0(t) \right\}, \tag{3.3}$$

where $S_\gamma^{(0)}(t, \mathbf{W}, \boldsymbol{\beta}_2) = \sum_{i=1}^m S_i(t_{ij}) e^{\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma \mathbf{W}_{ij}}$

Let

$$U_\gamma(\tau) = logL_\gamma = \sum_{i=1}^{m} \left\{ \int_0^\tau W(t_{ij}) dN_i(t) - \int_0^\tau W(t_{ij}) S(t_{ij}) e^{\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma \mathbf{W}_{ij}} dA_0(t) \right\} \tag{3.4}$$

and

$$\frac{\partial L_\gamma}{\partial \gamma} = U_\gamma(\tau)\ L_\gamma \left(\text{because } U_\gamma(\tau) = \frac{\partial log L\gamma}{\partial \gamma} = \frac{\partial L_\gamma}{\partial \gamma L_\gamma} \right). \tag{3.5}$$

The marginal likelihood of the longitudinal measurements is denoted by $l_1(\theta, Y)$ and the overall log likelihood is written as follows:

$l = l_1(\theta, Y) + logE_{W|Y}[L_\gamma]$

which the derivative with respect to γ is as follows:

$$\frac{\partial l}{\partial \gamma} = \frac{E_{W|Y}[U_\gamma(\tau) L_\gamma]}{E_{W|Y}[L_\gamma]} \tag{3.6}$$

The score statistic is shown as follows:

$$U_\gamma(\tau) = E_{W|Y}[U_0(\tau)]$$

$$= E_{W|Y}[\sum_{i=1}^{m} \left\{ \int_0^\tau W(t_{ij}) dN_i(t) - \int_0^\tau W(t_{ij}) S(t_{ij}) e^{\mathbf{x}'_{ij2}\boldsymbol{\beta}_2} dA_0(t) \right\}]$$

$$= \sum_{i=1}^{m} E_{W|Y}[W(t_{ij})] dM_i(t) \tag{3.7}$$

where $M_i(t) = N_i(t) - \Lambda_i(t) = N_i(t) - \int_0^{t_j} S_i(u) e^{\mathbf{x}'_{ij2}2} dA_0(u)$ is the usual counting process martingale. And $U_0(\tau)$ is $U_\gamma(\tau)$ under $\gamma = 0$. If U (t) is considered a particular value of a process $\{U(s) : s > 0\}$ and W is known it is predictable, then the variance of U (s) as follows:

$$V(s) = \sum_{i=1}^{m} \int_0^s E_{W|Y}[W(t_{ij})]^2 d\Lambda_i(t) \tag{3.8}$$

if the $W(t_{ij})$ between individuals are independent;

or

$$V(s) = \sum_{i=1}^{m} \left\{ \int_0^\tau E_{W|Y}[W(t_{ij})]^2 d\Lambda_i(t) - \int_0^\tau \int_0^\tau Cov_{W|Y}(W(t_{ij}), W(s_{ij})) dM_i(t) dM_i(s) \right\} \tag{3.9}$$

if the $W(t_{ij})$ between individuals are not independent.

According to the martingale central limit theorem under mild conditions, $U(s)/[V(s)]^{(1/2)}$ is asymptotically N (0, 1) under H_0 as $m \to \infty$.

3.3 LONGITUDINAL BIOMARKER FOR SURVIVAL

After the longitudinal biomarker is identified and fitted to the data, the fitted model is then considered to make inference about individuals survival at the future time t = t_2 given a previous time t = t_1. If $\mathbf{Y}_{i01}$ is a set of longitudinal measurements on the ith individual over the interval [0, t_1] then the survival function is wrritten as follows:

$S(t_2 \mid t_1, \mathbf{Y}_{i01}) = P(T > t_2 \mid T > t_1, \mathbf{Y}_{i01})$.

Evaluation of the conditional probability of surviving to t_2 involves the expectation with respect to the unobserved latent process. Let $\mathbf{W}_{i01}$ and $\mathbf{W}_{i02}$ be the values of $\mathbf{W}(t_{ij})$ within the intervals [0, t_1] and [0, t_2], respectively. Then $S(t_2 \mid t_1, \mathbf{Y}_{i01})$ can be written as

$$S(t_2 \mid t_1, \mathbf{Y}_{i01}) = \int P(T > t_2 \mid T > t_1, \mathbf{W}_{02}) f(\mathbf{W}_{02} \mid T > t_1, \mathbf{Y}_{i01}) d\mathbf{W}_{02}$$

$$= \frac{\int P(T > t_2 \mid T > t_1, \mathbf{W}_{02}) P(T > t_1 \mid \mathbf{W}_{02}) f(\mathbf{W}_{02} \mid \mathbf{Y}_{i01}) d\mathbf{W}_{02}}{P(T > t_1 \mid \mathbf{Y}_{i01})}$$

$$= \frac{\int P(T > t_2 \mid \mathbf{W}_{02}) f(\mathbf{Y}_{i01} \mid \mathbf{W}_{02}) f(\mathbf{W}_{02}) d\mathbf{W}_{02}}{\int P(T > t_1 \mid \mathbf{W}_{02}) f(\mathbf{Y}_{i01} \mid \mathbf{W}_{02}) f(\mathbf{W}_{02}) d\mathbf{W}_{02}}$$

$$= \frac{\int P(T > t_2 \mid \mathbf{W}_{02}) f(\mathbf{Y}_{i01} \mid \mathbf{W}_{01}) f(\mathbf{W}_{02}) d\mathbf{W}_{02}}{\int P(T > t_1 \mid \mathbf{W}_{01}) f(\mathbf{Y}_{i01} \mid \mathbf{W}_{01}) f(\mathbf{W}_{01}) d\mathbf{W}_{01}}. \tag{3.10}$$

If we ignore any information in $\mathbf{Y}_{i01}$, the conditional probability can be written as:

$$S(t_2 \mid t_1) = \frac{\int P(T > t_2 \mid \mathbf{W}_{02}) f(\mathbf{W}_{02}) d\mathbf{W}_{02}}{\int P(T > t_1 \mid \mathbf{W}_{01}) f(\mathbf{W}_{01}) d\mathbf{W}_{01}} \tag{3.11}$$

The term $f(\mathbf{Y}_{i01} \mid \mathbf{W}_{02})$ is a weighting factor which reflects the relevant information in $\mathbf{Y}_{i01}$. If it is possible to completely determine the true value $\mathbf{W}_{02}^0$ of $\mathbf{W}_{02}$ from $\mathbf{Y}_{i01}$, then $f(\mathbf{Y}_{i01} \mid \mathbf{W}_{01}) f(\mathbf{W}_{02})$ and $f(\mathbf{Y}_{i01} \mid \mathbf{W}_{01}) f(\mathbf{W}_{01})$ are zero under $\mathbf{W}_{02}^0 \neq \mathbf{W}_{02}$. Now we have maximum information from $\mathbf{Y}_{i01}$ and $S(t_2 \mid t_1, \mathbf{Y}_{i01}) = S(t_2 \mid t_1, \mathbf{W}_{02}^0)$ under $\mathbf{W}_{02}^0 = \mathbf{W}_{02}$.

3.4 MEASURING BIOMARKER EFFECTIVENESS

Two methods are considered to measure the effectiveness of a biomarker to predict survival at a future time, $t = t_2$, given the previous time $t = t_1$. One is a fixed point method and the other is an interval measures method. The key definition is described as follows: $S^0(t_{ij})$ is defined as the value of the observed survivor process for the i^{th} individual at time t_j. The value is one if the individual was known to be alive at t_j; the value is zero if the individual died before tj and the value is undefined if the individual was censored before t_j. If the biomarker is effective, there is the relatively small absolute deviation between $S^0(t_{ij})$ and the corresponding estimates $S(t_2 \mid t_1, \mathbf{Y}_{i01})$.

3.4.1 Fixed point method

In the fixed point method, the unbiased estimator for including the information of $\mathbf{Y}_{i01}$ is as follows:

$$M_Y(\tau_1, \tau_2) = \frac{1}{r(\tau_1)} \sum_{i:t_i \geq \tau_1} \Bigg[I(t_i \geq \tau_2)(1 - S(\tau_2 \mid \tau_1, \mathbf{Y}_{i01})) + \delta_i I(t_i < \tau_2)$$

$$S(\tau_2 \mid \tau_1, \mathbf{Y}_{i01}) + (1 - \delta_i) I(t_i < \tau_2) \Big\{ (1 - S(\tau_2 \mid \tau_1, \mathbf{Y}_{i01}))$$

$$\times S(\tau_2 \mid \tau_i, \mathbf{Y}_{i01}) + S(\tau_2 \mid \tau_1, \mathbf{Y}_{i01})(1 - S(\tau_2 \mid \tau_i, \mathbf{Y}_{i01})) \Big\} \Bigg], \tag{3.12}$$

where $r(\tau_1)$ is the number of individuals at risk at t_1 and δ_i is an indicator of censoring ($\delta_i = 0$) or observed failure ($\delta_i = 1$). And the unbiased estimator without the information of $\mathbf{Y}_{i01}$ as follows:

$$M(\tau_1, \tau_2) = \frac{1}{r(\tau_1)} \sum_{i:t_i \geq \tau_1} \Bigg[I(t_i \geq \tau_2)(1 - S(\tau_2 \mid \tau_1)) + \delta_i I(t_i < \tau_2)$$

$$S(\tau_2 \mid \tau_1) + (1 - \delta_i) I(t_i < \tau_2) \Big\{ (1 - S(\tau_2 \mid \tau_1))$$

$$\times S(\tau_2 \mid \tau_i) + S(\tau_2 \mid \tau_1)(1 - S(\tau_2 \mid \tau_i)) \Big\} \Bigg], \tag{3.13}$$

A relative measure can be used to interpret the effectiveness of the biomarker by comparison of $M_Y(\tau_1, \tau_2)$ and $M(\tau_1, \tau_2)$ as follows:

$$R_M(\tau_1, \tau_2) = 1 - M_Y(\tau_1, \tau_2)/M(\tau_1, \tau_2). \tag{3.14}$$

3.4.2 Interval measures method

An alternative procedure for measuring the effectiveness of a biomarker is the interval measures method. The unbiased estimator for including the information of $\mathbf{Y}_{i01}$ for this method is:

$$D_Y(\tau_1, \tau_2) = \frac{\sum_{i:\tau_1 \leq t_i \leq \tau_2} \delta_i \hat{G}(\tau_1)/\hat{G}(t_i) M_Y(\tau_1, t_i)}{\sum_{i:\tau_1 \leq t_i \leq \tau_2} \delta_i \hat{G}(\tau_1)/\hat{G}(t_i)} \tag{3.15}$$

where $\hat{G}(\cdot)$ is the Kaplan-Meier estimator of the censoring time distribution, which is used to compensate for the loss of censored cases. The unbiased estimator without the information of $\mathbf{Y}_{i01}$ is:

$$D(\tau_1, \tau_2) = \frac{\sum_{i:\tau_1 \leq t_i \leq \tau_2} \delta_i \hat{G}(\tau_1)/\hat{G}(t_i) M(\tau_1, t_i)}{\sum_{i:\tau_1 \leq t_i \leq \tau_2} \delta_i \hat{G}(\tau_1)/\hat{G}(t_i)} \tag{3.16}$$

A relative measure can be used to interpret the effectiveness of the biomarker by comparison of $M_Y(\tau_1, \tau_2)$ and $M(\tau_1, \tau_2)$ as follows:

$$R_D(\tau_1, \tau_2) = 1 - D_Y(\tau_1, \tau_2)/D(\tau_1, \tau_2). \tag{3.17}$$

3.5 THE COMBINATION OF SURVIVAL WITH HETEROGENEOUS HAZARD ANALYSIS AND LONGITUDINAL DATA ANALYSIS

To formulate our model, we recognize that different individuals in a population can have vastly different underlying risks of having an event of interest. Consequently, a *frailty* model can be used to extend the proportional hazards regression model in survival analysis. Similar to the formulation of the frailty model provided by Klein and Moeschberger (1997) [23], we can write the hazard rate at time t in i^{th} patient as

$$h_i(t) = h_0(t)\exp(\alpha\kappa_i + \boldsymbol{\beta}^t\mathbf{x}_{ij}), i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n_i \tag{3.18}$$

where $h_0(t)$ is an arbitrary baseline hazard rate, $\mathbf{x_{ij}}$ is the vector of covariates, $\boldsymbol{\beta}$ is the vector of regression coefficients, and $\kappa_1, \ldots, \kappa_m$ are the frailties. It is usually assumed that the κ's consist of an independent sample from some distribution with mean 0 and variance 1. If α is zero, then the above equation reduces to Cox's proportional hazards model. A more convenient form of model (3.18) can be written as

$$h_i(t) = h_0(t)q_i\exp(\boldsymbol{\beta}^t\mathbf{x_{ij}}),\ i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n_i \tag{3.19}$$

Turning to the longitudinal part of the model, if we let $\mathbf{Y}_i(t)$ represent the underlying biomarker vector of the i^{th} individual at time t, then, following Henderson, et al., 2002, we can write

$$\mathbf{Y}_i(t) = \mathbf{x}'_{i1}(t)\boldsymbol{\beta}_1 + \mathbf{W}_i(t) + \mathbf{e}_i(t). \tag{3.20}$$

where $\mathbf{x}_{i1}(t)$ is a $p_1 \times 1$ vector of explanatory variables, and $\mathbf{W}_i(t)$ and $\mathbf{e}_i(t)$ are zero-mean random processes. Since the biomarkers are sampled at discrete time points and assuming

that the $\mathbf{Y}_i$ are from a normal distribution, a discrete version of the model can be re–written as

$$\mathbf{Y}_{ij} = \mathbf{x}'_{ij1}\boldsymbol{\beta}_1 + \mathbf{W}_{ij} + \mathbf{e}_{ij}, \quad i = 1, 2, \ldots, m, \ j = 1, 2, \ldots, n_i \ , \tag{3.21}$$

where $\mathbf{W}_{ij}$ is the value of a zero-mean Gaussian random effect for the i^{th} individual at the j^{th} time point, $\mathbf{e}_{ij}$ is a zero-mean Gaussian measurement error and n_i is the number of observations for individual i. The j's in the discrete version of the model refer to the last biomarkers observed at or before time t. The errors, $\mathbf{e}_{ij}$, are assumed here to be mutually independent and the within individual correlation in $\mathbf{Y}_{ij}$ arises through serial correlation in the random effect, $\mathbf{W}_{ij}$.

In a frailty model extension of the proportional hazards regression introduced by Henderson, et al., 2002, the probability density function is

$$f(t) = S(t)\lambda_0(t)q_i \exp\{\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma\mathbf{W}_{ij}\} \tag{3.22}$$

where $S(t)$ is a predictable survival function, $\lambda_0(t)$ is an unspecified baseline hazard, q_i is the unobervable frailty from independent and identically distributed sample of gamma random variables, and $\mathbf{x}_{ij2}$ is a $p_2 \times 1$ vector of explanatory variables. As indicated by equation (3.22), the connection between the longitudinal process and the failure process is made through the parameter associated with the latent process, $\mathbf{W}_{ij}$.

For q_i, the probabilty density function is

$$g(q) = \frac{q^{(1/\nu - 1)} \exp(-q/\nu)}{\Gamma[1/\nu]\nu^{1/\nu}} \ . \tag{3.23}$$

If $\mathbf{W}$ is known, the conditional likelihood of the survival data over all times and individuals can be written as follows:

$$L_\gamma = \left(\prod_i \prod_j \left(e^{\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma\mathbf{W}_{ij}} \widehat{q}_i dA_0(t_{ij}) \right)^{\Delta N(t_{ij})} \right) \exp\left\{ - \int_0^\tau S_\gamma^{(0)}(t_{ij}, W, \boldsymbol{\beta}_2) dA_0(t_{ij}) \right\} \tag{3.24}$$

where $S_\gamma^{(0)}(t_{ij}, \mathbf{W}, \boldsymbol{\beta}_2) = \sum_{i=1}^m \widehat{q}_i S_i(t_{ij}) e^{\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma \mathbf{W}_{ij}}$ and $A_0(t_{ij})$ is the cumulative baseline intensity.

The likelihood of the survival data associated with the frailty conditional on the latent process, $\mathbf{W}$ can be written as follows:

$$L_{\text{FULL}} = L_\nu \times L_\gamma$$

$$= \prod_i \left(\frac{q_i^{(1/\nu - 1)} \exp(-q_i/\nu)}{\Gamma[1/\nu]\nu^{1/\nu}} \prod_j \left(e^{\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma \mathbf{W}_{ij}} \widehat{q}_i dA_0(t_{ij}) \right)^{\Delta N(t_{ij})} \right) \exp\left\{ - \int_0^\tau S_\gamma^{(0)}(t_{ij}, W, \boldsymbol{\beta}_2) dA_0(t_{ij}) \right\} \quad (3.25)$$

The partial conditional likelihood of the survival data associated with the frailty can then be written as the sum of the log likehood associated with the frailty distribution plus that associated with the Cox regresssion model, that is,

$$\ell_{\text{FULL}} = \ell_\nu + \ell_\gamma(\boldsymbol{\beta}, \gamma, A_0),$$

where

$$\ell_\nu = -m \left[(1/\nu) \ln \nu + \ln \Gamma[1/\nu] \right] + \sum_{i=1}^m \left\{ [1/\nu - 1] \ln q_i - q_i/\nu \right\}, \quad (3.26)$$

and

$$\ell_\gamma(\boldsymbol{\beta}_2, \gamma, A_0) = \sum_{i=1}^m \sum_{j=1}^{n_i} \delta_{ij} [\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma \mathbf{W}_{ij} + \ln dA_0(t_{ij})] - q_i A_0(t_{ij}) \exp(\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma \mathbf{W}_{ij}) \quad (3.27)$$

where δ_{ij} is 1 if individual i has an event at the j^{th} time point and 0, otherwise.

The EM algorithm provides a means of maximizing complex likelihoods. Here we use EM algorithm described by Klein and Moeschberger (2003) [57]. In the E–step of the algorithm, the expected value of ℓ_{FULL} is computed, given the current estimates of the parameters and the observable data. In the M–step of the algorithm, estimates of the parameters which maximize the expected value of ℓ_{FULL} from the E–step are obtained. The algorithm iterates between these two steps until convergence.

To apply the E–step, we'll assume, similar to Klein and Moeschberger (2003) [24], that the $q_i's$ are independent gamma random variables with shape parameters $B_i = 1/\nu + \sum_{j=1}^{n_i} \delta_{ij}$, and scale parameters, $C_i = 1/\nu + \sum_{i=1}^{m} A_0(t_{ij}) \exp(\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma \mathbf{W}_{ij})$. Thus,

$$E[q_i \mid \text{Data}] = \frac{B_i}{C_i} \quad \text{and} \quad E[\ln q_i] = \psi(B_i) - \ln C_i, \tag{3.28}$$

where $\psi(\cdot)$ is the digamma function. Substituting these values in (3.26) and (3.27) completes the E–step of the algorithm.

For the M–step, $E[\ell_\gamma(\boldsymbol{\beta}_2, \gamma, A_0) \mid \text{Data}]$ is expressed as

$$\ell_\gamma(\boldsymbol{\beta}_2, \gamma, A_0) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \delta_{ij}[(\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma \mathbf{W}_{ij}) + \ln dA_0(t_{ij})] - \frac{B_i}{C_i} A_0(t_{ij}) \exp(\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma \mathbf{W}_{ij}) \tag{3.29}$$

The expression in equation (3.29) is associated with the nuisance parameter $dA_0()$.

Now, let $t_{(k)}$ be the k^{th} smallest event time, and $b_{(k)}$ be the number of events at time $t_{(k)}$, $k = 1, \ldots, F$. Also, denote the expected value of the frailty and the covariate vector for the h^{th} individual in the risk set $R(t_{(k)})$ by $\widehat{q}_h$ and, $\mathbf{x}_h$ and $\mathbf{W}_h$, respectively. Then, the partial likelihood to be maximized in the M–step is

$$\ell_\gamma(\boldsymbol{\beta}_2, \gamma) = \sum_{k=1}^{F} \left\{ S_{(k)} - b_{(k)} \ln \left[\sum_{h \in R(t_{(k)})} \hat{q}_h \exp(\mathbf{x_h}'\boldsymbol{\beta}_2 + \gamma \mathbf{W}_h) \right] \right\} , \tag{3.30}$$

where $S_{(k)}$ is the sum of the covariates of individuals who had an event at time $t_{(k)}$.

An estimate of $A_0(t_{ij})$ from this step is given by

$$\hat{A}_0(t_{ij}) = \sum_{t_{(k)} \le t_{ij}} dA_{k_0} \tag{3.31}$$

where

$$dA_{k_0} = \frac{b_{(k)}}{\sum_{h \in R(t_{(k)})} \hat{q}_h \exp(\mathbf{x_h}'\boldsymbol{\beta}_2 + \gamma \mathbf{W}_h)}$$

A full implementation of the EM algorithm is as follows (Klein and Moeschberger [2003]):

Step 0 Provide initial estimates of $\boldsymbol{\beta}_2, \gamma, \nu$ and thus h_{k_0}, $k = 1, \ldots, F$.

Step 1 (E–step) Compute B_i, C_i, $i = 1, \ldots, m$ and $\hat{q}_h$, $h = 1, \ldots, n_i$ based on the current values of the parameters.

Step 2 (M–step) Update the estimate of $\boldsymbol{\beta}_2, \gamma$ (and the dA_{k_0}) using the partial likelihood. Update the estimate of ν based on the likelihood $\ell_{\nu|\text{Data}} = E[\ell_\nu \mid \text{Data}]$ given by

$$\ell_{\nu|\text{Data}} = -m\left[(1/\nu)\ln\nu + \ln\Gamma(1/\nu)\right] + \sum_{i=1}^{m}\left\{[1/\nu - 1][\psi(B_i) - \ln C_i] - \frac{B_i}{\nu C_i}\right\}$$

Step 3 Iterate between Steps 1 and 2 until convergence.

Now, let

$$U_\gamma(\tau) = logL_\gamma = \sum_{i=1}^{m}\left\{\int_0^\tau W(t_{ij})dN(t_{ij}) - \int_0^\tau W(t_{ij})\hat{q}_i S(t_{ij}) e^{\mathbf{x}'_{ij2}\boldsymbol{\beta}_2 + \gamma \mathbf{W}_{ij}} dA_0(t_{ij})\right\} \quad (3.32)$$

and

$$\frac{\partial L_\gamma}{\partial \gamma} = U_\gamma(\tau)\ L_\gamma\left(\text{because } U_\gamma(\tau) = \frac{\partial logL\gamma}{\partial \gamma} = \frac{\partial L_\gamma}{\partial \gamma L_\gamma}\right) . \quad (3.33)$$

The resulting score statistic is

$$U_\gamma(\tau) = E_{W|Y}[U_0(\tau)]$$

$$= E_{W|Y}\left[\sum_{i=1}^{m}\left\{\int_0^\tau W(t_{ij})dN(t_{ij}) - \int_0^\tau W(t_{ij})\hat{q}_i S(t_{ij}) e^{\mathbf{x}'_{ij2}\boldsymbol{\beta}_2} dA_0(t_{ij})\right\}\right]$$

$$= \sum_{i=1}^{m} E_{W|Y}[W(t_{ij})]dM(t_{ij}) , \quad (3.34)$$

where $M(t_{ij}) = N(t_{ij}) - \Lambda(t_{ij}) = N(t_{ij}) - \int_0^{t_{ij}} \hat{q}_i S_i(u) e^{\mathbf{x}'_{ij2}\boldsymbol{\beta}_2} dA_0(u)$ is the usual counting process martingale (see Fleming and Harrington (1991) [25]) and $U_0(\tau) = U_\gamma(\tau)$ if $\gamma = 0$.

We consider $U(t)$ to be a particular value of a process, $\{U(s) : s > 0\}$ and W to be known and predictable so that the variance of $U(s)$ is

$$V(s) = \sum_{i=1}^{m} \int_0^s E_{W|Y}[W(t_{ij})]^2 d\Lambda(t_{ij}), \tag{3.35}$$

if the $W(t_{ij})$ between individuals are independent; or

$$V(s) = \sum_{i=1}^{m} \left\{ \int_0^\tau E_{W|Y}[W(t_{ij})]^2 d\Lambda(t_{ij}) - \int_0^\tau \int_0^\tau Cov_{W|Y}(W(t_{ij}), W(s_{ij}))dM(t_{ij})dM(s_{ij}) \right\} \tag{3.36}$$

if the $W(t_{ij})$ between individuals are not independent.

According to the martingale central limit theorem under mild conditions, $U(s)/[V(s)]^{(1/2)}$ is asymptotically N (0, 1) under H_0 as $m \to \infty$.

3.6 LONGITUDINAL BIOMARKER FOR SURVIVAL WITH HETEROGENEOUS HAZARD SURVIVAL

In Section 3.3, we showed how a longitudinal biomarker is identified and fitted to data. Next, the fitted model is used to make inference about individuals' survival at a future time t = t_2 given an earlier time t = t_1. In this section, we extend the model to accommodate frailty in the survival data. Let $\mathbf{Y}_{i01}$ be longitudinal measurements on the ith individual over the interval [0, t_1]. Then the survival function is as follows:

$S(t_2 \mid t_1, \mathbf{Y}_{i01}, \kappa) = P(T > t_2 \mid T > t_1, \mathbf{Y}_{i01}, \kappa)$.

This is the same survival function as in section 3.3 but with the addition of a frailty parameter κ.

Evaluation of the conditional probability of surviving to t_2 involves the expectation with respect to the unobserved latent process. Let $\mathbf{W}_{i01}$ and $\mathbf{W}_{i02}$ be the values of $\mathbf{W}(t_{ij})$ at measurement within the intervals [0, t_1] and [0, t_2], respectively. Then $S(t_2 \mid t_1, \mathbf{Y}_{i01}, \kappa)$ can be written the form as follows:

$$S(t_2 \mid t_1, \mathbf{Y}_{i01}, \kappa) = \int P(T > t_2 \mid T > t_1, \mathbf{W}_{02}, \kappa) f(\mathbf{W}_{02} \mid T > t_1, \mathbf{Y}_{i01}) d\mathbf{W}_{02}$$

$$= \frac{\int P(T > t_2 \mid T > t_1, \mathbf{W}_{02}, \kappa) P(T > t_1 \mid \mathbf{W}_{02}, \kappa) f(\mathbf{W}_{02} \mid \mathbf{Y}_{i01}) d\mathbf{W}_{02}}{P(T > t_1 \mid \mathbf{Y}_{i01}, \kappa)}$$

$$= \frac{\int P(T > t_2 \mid \mathbf{W}_{02}, \kappa) f(\mathbf{Y}_{i01} \mid \mathbf{W}_{02}) f(\mathbf{W}_{02}) d\mathbf{W}_{02}}{\int P(T > t_1 \mid \mathbf{W}_{02}, \kappa) f(\mathbf{Y}_{i01} \mid \mathbf{W}_{02}) f(\mathbf{W}_{02}) d\mathbf{W}_{02}}$$

$$= \frac{\int P(T > t_2 \mid \mathbf{W}_{02}, \kappa) f(\mathbf{Y}_{i01} \mid \mathbf{W}_{01}) f(\mathbf{W}_{02}) d\mathbf{W}_{02}}{\int P(T > t_1 \mid \mathbf{W}_{01}, \kappa) f(\mathbf{Y}_{i01} \mid \mathbf{W}_{01}) f(\mathbf{W}_{01}) d\mathbf{W}_{01}} \tag{3.37}$$

If we ignore any information in $\mathbf{Y}_{i01}$, the conditional probability can be written as:

$$S(t_2 \mid t_1, \kappa) = \frac{\int P(T > t_2 \mid \mathbf{W}_{02}, \kappa) f(\mathbf{W}_{02}) d\mathbf{W}_{02}}{\int P(T > t_1 \mid \mathbf{W}_{01}, \kappa) f(\mathbf{W}_{01}) d\mathbf{W}_{01}} \tag{3.38}$$

The term $f(\mathbf{Y}_{i01} \mid \mathbf{W}_{01})$ is a weighting factor which reflects the relevant information in $\mathbf{Y}_{i01}$. If it is possible to completely determine the true value $\mathbf{W}_{02}^0$ of $\mathbf{W}_{02}$ from $\mathbf{Y}_{i01}$, then $f(\mathbf{Y}_{i01} \mid \mathbf{W}_{01}) f(\mathbf{W}_{02})$ and $f(\mathbf{Y}_{i01} \mid \mathbf{W}_{01}) f(\mathbf{W}_{01})$ are zero if $\mathbf{W}_{02}^0 \neq \mathbf{W}_{02}$. Now we have maximum information from $\mathbf{Y}_{i01}$ and $S(t_2 \mid t_1, \mathbf{Y}_{i01}, \kappa) = S(t_2 \mid t_1, \mathbf{W}_{02}^0, \kappa)$ if $\mathbf{W}_{02}^0 = \mathbf{W}_{02}$.

3.7 MEASURING BIOMARKER EFFECTIVENESS AT THE SURVIVAL WITH HETEROGENEOUS HAZARD ANALYSIS

In Section 3.4, two methods were considered to measure the effectiveness of tumor size to predict survival at time $t = t_2$ given $t = t_1$. One is a fixed point method and the other was an interval measures method. In this section, both methods will be extended to again accommodate the multivariate survival data. Again, a frailty model is chosen for the multivariate data. The key definition is described as follows: $S^0(t_{ij})$ is defined as the value of the observed survivor process for the i^{th} individual at time t_j. The value is one if the individual was known to be alive at t_j; the value is zero if the individual died before t_j and the value is undefined if the individual was censored before t_j. If the biomarker is effective, there is a relatively small absolute deviation between $S^0(t_{ij})$ and the corresponding estimates $S(t_2 \mid t_1, \mathbf{Y}_{i01}, \kappa)$.

3.7.1 Fixed point method

In the fixed point method, the unbiased estimator for including the information of $\mathbf{Y}_{i01}$ is as follows:

$$M_{Y_1}(\tau_1,\tau_2)=\frac{1}{r(\tau_1)}\sum_{i:t_i\geq\tau_1}\Bigg[I(t_i\geq\tau_2)(1-S(\tau_2\mid\tau_1,\mathbf{Y_{i01}},\kappa))+\delta_i I(t_i<\tau_2)$$

$$S(\tau_2\mid\tau_1,\mathbf{Y}_{i01},\kappa)+(1-\delta_i)I(t_i<\tau_2)\Big\{(1-S(\tau_2\mid\tau_1,\mathbf{Y}_{i01},\kappa))$$

$$\times S(\tau_2\mid\tau_i,\mathbf{Y}_{i01},\kappa)+S(\tau_2\mid\tau_1,\mathbf{Y}_{i01},\kappa)(1-S(\tau_2\mid\tau_i,\mathbf{Y}_{i01},\kappa))\Big\}\Bigg], \tag{3.39}$$

where $r(\tau_1)$ is the number of individuals at risk at t_1 and δ_i is an indicator of censoring ($\delta_i = 0$) or observed failure ($\delta_i = 1$). And the unbiased estimator without the information of $\mathbf{Y}_{i01}$ as follows:

$$M_1(\tau_1,\tau_2)=\frac{1}{r(\tau_1)}\sum_{i:t_i\geq\tau_1}\Bigg[I(t_i\geq\tau_2)(1-S(\tau_2\mid\tau_1,\kappa))+\delta_i I(t_i<\tau_2)$$

$$S(\tau_2\mid\tau_1,\kappa)+(1-\delta_i)I(t_i<\tau_2)\Big\{(1-S(\tau_2\mid\tau_1,\kappa))$$

$$\times S(\tau_2\mid\tau_i,\kappa)+S(\tau_2\mid\tau_1,\kappa)(1-S(\tau_2\mid\tau_i,\kappa))\Big\}\Bigg]. \tag{3.40}$$

A relative measure can be used to interpret the effectiveness of the biomarker by comparison of $M_{Y_1}(\tau_1,\tau_2)$ and $M_1(\tau_1,\tau_2)$ as follows:

$$R_{M_1}(\tau_1,\tau_2)=1-M_{Y_1}(\tau_1,\tau_2)/M_1(\tau_1,\tau_2). \tag{3.41}$$

3.7.2 Interval measures method

The alternative procedure for measuring the effectiveness of a biomarker is the interval measures method. The unbiased estimator for including the information of $\mathbf{Y}_{i01}$ for this method is:

$$D_{Y_1}(\tau_1, \tau_2) = \frac{\sum_{i:\tau_1 \leq t_i \leq \tau_2} \delta_i \hat{G}(\tau_1)/\hat{G}(t_i) M_{Y_1}(\tau_1, t_i)}{\sum_{i:\tau_1 \leq t_i \leq \tau_2} \delta_i \hat{G}(\tau_1)/\hat{G}(t_i)} \tag{3.42}$$

where $\hat{G}(\cdot)$ is the Kaplan-Meier estimator of the censoring time distribution, which is used to compensate for the loss of censored cases. The unbiased estimator without the information of $\mathbf{Y_{i01}}$ is:

$$D_1(\tau_1, \tau_2) = \frac{\sum_{i:\tau_1 \leq t_i \leq \tau_2} \delta_i \hat{G}(\tau_1)/\hat{G}(t_i) M_1(\tau_1, t_i)}{\sum_{i:\tau_1 \leq t_i \leq \tau_2} \delta_i \hat{G}(\tau_1)/\hat{G}(t_i)} \tag{3.43}$$

A relative measure can be used to interpret the effectiveness of the biomarker by comparison of $M_{Y_1}(\tau_1, \tau_2)$ and $M_1(\tau_1, \tau_2)$ as follows:

$$R_{D_1}(\tau_1, \tau_2) = 1 - D_{Y_1}(\tau_1, \tau_2)/D_1(\tau_1, \tau_2). \tag{3.44}$$

Equations (3.37) through (3.44) are the same as equations (3.10) through (3.17) except the latter set of equations includes a frailty parameter, κ.

4.0 RESULTS

4.1 SIMULATION STUDY

4.1.1 Power of score test for association between longitudinal biomarker values and survival time function

In order to investigate the empirical properties of the score test for association between longitudinal biomarker values and the survival time function, we performed a simulation study which was similar in strategy to that used by Henderson, et al. (2002). However, in their paper, $E[Y]$ was a linear function of time and the survival function was $\exp(-0.1t^2)$, that is, the failure times were from a Weibull distribution. In our simulation, both $E[Y]$ and the survival function are from gamma distributions. We also constructed the survival function to reflect that every subject has his/her own frailty. We examined the empirical type I error rates of the score tests, that is, the power under $H_0 : \gamma = 0$. Other alternative hypotheses were also explored.

Sample sizes were constructed as follows: total sample size = number of subjects $\times$ number of observation times or $N = n \times T$. The total number of observations, N, within a given set of simulations was fixed at a constant number but the number of subjects, n, and the number of time points, T were varied accordingly. The time points, t, were chosen to be integer values between 1 and T, inclusively. In each simulation, the frailty parameter varies across subjects for each simulation. The EM algorithm used for the simulations is provided

in the coxph function in S-plus 6.2. One thousand (1000) realizations were generated for each sample size.

In the first set of simulations, we modeled the generated survival data by (correctly) using frailty parameters to characterize the individuals' heterogeneous baseline hazards. The simulation results are shown in Tables 4-1A – 4-1C. Three different latent process types were specified for assessing the power of score test. The structures for the three different latent process types are as follows:

$$(\mathbf{1}) : W(t) = U_1,\ U_1 \sim N(0, \sigma_1^2)$$

$$(\mathbf{2}) : W(t) = U_1 + U_2 \times t,\ U_1 \sim N(0, \sigma_1^2),$$

$$U_2 \sim N(0, \sigma_2^2),\ \mathrm{Corr}(U_1, U_2) = \rho$$

$$(\mathbf{3}) : W(t) = U_1 + V(t),\ U_1 \sim N(0, \sigma_1^2),\ V(t) \sim N(0, \sigma_2^2),$$

$$\mathrm{Corr}(V(t),\ V(t+s)) = \exp(-|s|)$$

We also examined the power of the score test under $H_0 : \gamma = 0$ while ignoring the existence of the underlying frailty. Similar situations to those described above were considered except the frailty was now ignored in the models. Results are shown in Tables 4-2A – 4-2C.

From the results of Tables 4–1 and 4–2, some empirical properties are as follows:

(1) For latent process types (2) and (3), a higher correlation between longitudinal biomarker values and survival time function given the large values of $\sigma_1^2 + \sigma_2^2$ results in higher values for the power of the score test.

(2) For latent process type (1), there are not substantial differences in the power of the score test for different values of σ_1^2 when the sample sizes are small but when the sample sizes are large, then for all latent process types, larger values of σ_1^2 result in higher powers associated with the score test.

(3) For latent process type (2), there are not substantial differences in the power of the score test for different values of ρ between σ_1^2 and σ_2^2.

By comparing Tables 4–1 and 4–2, one concludes that the power of the score test in models which correctly specify the existence of the frailty are higher than models that ignore the existence of the frailty when survival data with heterogeneous hazard structures were modeled.

4.1.2 Examination of the measure of biomarker effectiveness in a survival outcomes with heterogeneous hazard functions

In order to investigate some empirical properties for the measure of biomarker effectiveness in survival analyses with the presence of heterogeneous hazards among individuals, we conducted more simulations. First, we varied the sample sizes from 25 to 1000 subjects and varied the number of time points at which the longitudinal biomarkers were measured from 2 to 20. We considered four levels of censoring for the time to event outcome: no censoring, low censoring (10%), medium censoring (25%) and high censoring (50%).

For this set of simulations, we only considered latent process type (3) as described in section 4.1.1. Simulation results for the fixed point method (described in section 3.7.1) are shown in Table 4–3. Simulation results for the interval measure method (described in section 3.7.2) are shown as Table 4–4. Each entry in Tables 4–3 and 4–4 was based on 1000 realizations of modeling the particular combination of sample size, number of time points and censoring level.

4.1.3 Power of the score test for association between longitudinal biomarkers and survival time when the biomarkers are missing at random

Since the modeling of missing values for longitudinal biomarkers is an important issue, we also conducted a simulation study to examine its effect on our models. First, we compared our method with the method of Henderson, Diggle, and Dobson under missing at random for longitudinal biomarker values, and then we compared our method with the method of Henderson, Diggle, and Dobson under nonignorable missing for longitudinal biomarker values. Three different latent process types were specified for assessing the power of score test. The structures for the three different latent process types are as section 4.1.1 described. The percentage of missing biomarkers is 50 %. Again, the results of the simulation were based on 1000 realizations of each scenario.

Simulation results under missing at random for longitudinal biomarker values when accounting for the frailty structure are shown in Tables 4-5A – 4-5C. Results of analogous simulations but while ignoring the frailty structure are shown in Tables 4-6A – 4-6C.

Comparisons of Tables 4–5 and 4–6 indicate that the power of the score test one specifies the existence of the frailty structure is higher than that when one ignores the existence of the frailty .

4.1.4 Power of the score test for association between longitudinal biomarkers and survival time when the missingness in the biomarkers is nonignorable

We also considered situations where the missingness in the biomarker values was nonignorable. We examined the same situations as for the missing at random cases; namely, 1) correctly considering the frailty structure and 2) incorrectly ignoring the frailty structure. Simulation results for situation 1) are shown in Tables 4-7A – 4-7C and for situation 2) are shown in Tables 4-8A – 4-8C.

Comparisons of Tables 4–7 and 4–8 indicate that the power of the score test is not good when the missingness of the longitudinal biomarkers is nonignorable *regardless* of whether or not the frailty structure is accounted for. However, correctly considering the existence of the frailty in these models is still better than ignoring its existence.

4.2 APPLICATION OF EXTENSIONS TO ANALYSIS OF SURVIVAL IN LIVER CIRRHOSIS PATIENTS

We next consider a dataset from a randomized trial in liver cirrhosis (Andersen, et al. (1993)). This data was also analyzed by Henderson et al. (2002) and is available on the web site, http://staff.pubhealth.ku.dk/˜pka/.

4.2.1 Description of the data and the data analysis

There were 488 patients in the trial, 251 patients of whom were assigned into the treatment (Prednisone) group and 237 patients of whom were assigned into the placebo group. All patients were followed until death or end of study. The dataset contained several variables: patient id, treatment status, current prothrombin value, current measurement time, previous prothrombin value, previous measurement time and a censoring indicator. We focus in this dissertation on the prothrombin biomarker which is measured repeatedly and on the treatment variable as they relate to the overall survival in these patients. In the first part of the analysis, the latent process was considered to be a random effect and was calculated by using a mixed model approach. Next, the score test for the $W(t)$ model under $H_0 : \gamma = 0$ was calculated using a Cox frailty model. The results are shown in Table 4–9. Then the marker effectiveness for prothrombin was determined, the result was shown in Table 4-10. and Table 4-11. as follows:

4.2.2 Discussion of our extensions of survival as applied to the liver cirrhosis dataset

From Table 4–9, it can be seen that whether considering the frailty or not, the score test for the longitudinal biomarker, prothrombin, under all of three latent process models are significant. Thus, some would consider this biomarker as a surrogate of survival for patients who have liver cirrhosis. Comparisons of the results of log likelihood ratio tests from the three different types of latent models indicate that latent process types (2) and (3) have larger test values than latent process type (1). Hence, latent process types (2) and (3) are more suitable than latent type (1) for these data.

From Tables 4–10 and 4–11, one can make conclusions about both the early (Year 0 to Year 1) and late (Year 3 to Year 4) effectiveness of the prothrombin biomarker as a surrogate for survival. At early times, the effectiveness of prothrombin is very significant when frailty is included in the model. In the fixed point method, R_{M_1} for latent process type (1) is 0.243 and for both latent process types (2) and (3) is 0.310. For the interval measure method, R_{D_1} is 0.738 for latent process type (1) and for latent process types (2) and (3) is 0.768. Even when frailty is not included in the model, the early effectiveness of prothrombin (at times 0 to year 1) is very significant in these liver cirrhosis patients. In fixed point method, R_{M_1} for latent process type (1) is 0.038 and for latent process type (2) and type (3) is 0.294. In interval measure method, R_{D_1} for latent process type (1) is 0.690 and for latent process type (2) and type (3) is 0.731.

At late times, the effectiveness of prothrombin was not significant regardless of whether or not frailty was included in the model. Results also did not change based on whether or not the fixed point method or the interval measure method was used and was not dependent on the type of latent process model being used. In all cases, $R_{M_1} = 0$. Thus, the effectiveness of the biomarker appears to diminish at late times.

5.0 DISCUSSION

5.1 DISCUSSION OF OUR EXTENSION

In the Henderson et al paper (2002), they focused only on random effects in the longitudinal biomarker. They put random effects in the Cox model to determine if longitudinal biomarker is associated with survival. However, in our method, we deal with random effects from the longitudinal biomarker and survival. We extended Henderson et al method to deal with survival with heterogeneous hazard. In our method, we can determine if longitudinal biomarker is associated with survival while simultaneously considering the existence of frailty in the survival.

Another aspect of our study was the fact that we estimated the influence of missing data for power of score test for association between longitudinal biomarker values and survival time function and biomarker effectiveness. Because the missing data is an important problem when we analyze the biological and clinical data, we want to know if our method is better than Henderson et al method under the existence of frailty in the survival.

5.2 CONCLUSIONS FROM SIMULATION STUDY AND APPLICATION OF EXTENSIONS OF SURVIVAL IN LIVER CIRRHOSIS PATIENTS

5.2.1 Conclusions from the Simulation Studies

5.2.1.1 Power of score test for association between longitudinal biomarker values and survival time function For large sample sizes, our model is better for detecting as a longitudinal biomarker as a predictor for survival data than the method used in the Henderson et al (2002) paper. However, for small sample sizes, our method also results in a large overinflation of the α-level as compared to the method using in the Henderson et al paper. An obvious conclusion is that a larger sample size for the survival data is associated with a larger power to detect a longitudinal biomarker as a surrogate for survival data than the small sample size for the survival data.

Also the longitudinal biomarker can be more easily identified as a surrogate for survival when the random effects from the longitudinal biomarker are either latent process types (2) or (3). A possible reason is that the latent process types (2) and (3) can show that the change of random effects is associated with time so they can be adequate to describe survival situation on an individual level. On the other hand, the latent process type (1) can show that the change of random effects is not associated with time so it is worse to describe survival situation in the individual level than the latent process types (2) and (3).

5.2.1.2 Measuring biomarker effectiveness for predicting survival with heterogeneous hazard analysis The interval measures method is a modified method for the fixed point method. The weighting factor for the interval measures method is the numbers of uncensoring observations between t_1 and t_2. If the percentage of censoring is high, the interval measures method is better than the fixed point method if censoring is MCAR.

Given a particular sample size and considering the dependency of the survival data, an adequate study design to determine if longitudinal biomarkers are surrogates for survival data to have small numbers of subjects and large numbers of observed time points. Although

relatively large numbers of subjects and small numbers of observed time points have a relative large power, assessing biomarker effectiveness is relatively smaller than for small numbers of subjects and large numbers of observed time points.

5.2.1.3 Power of the score test for association between longitudinal biomarkers and survival time when the biomarkers are missing at random The power of the score test under missing at random for the longitudinal biomarker is similar as the one for the complete data though the power under missing at random for the longitudinal biomarker is less than the complete data. However, our method appears to be better than the method of Henderson, Diggle, and Dobson under missing at random for detecting a longitudinal biomarker when frailty exists in the survival endpoint.

5.2.1.4 Power of the score test for association between longitudinal biomarkers and survival time when the missingness in the biomarkers is nonignorable The power of the score test under nonignorable missing shows that the results arsubstantially worse than the complete data. There is also an even greater overinflation of the α-level. However, our method appears to be better than the method of Henderson, Diggle, and Dobson if frailty exists in the hazard. However, our method cannot distinguish low association from no association between survival time and the longitudinal biomarkers if the missingness in the longitudinal biomarker is nonignorable. However, it can work on the data under nonignorable missing for the longitudinal biomarker if there is medium or high association between survival time and the longitudinal biomarker.

5.3 FURTHUR DIRECTIONS

Our method had several shortfalls. One was the large overinflation of the α-level, that is, the power under H_0, for the score tests when the sample sizes were small. Such an undesirable

property requires one to consider modifications to our large sample approximations of the score test to ensure that the α-levels would be closer to nominal values.

Another improvement to the frailty method proposed here would be to combine our method and methods that deal with both ignorable and nonignorable missing data mechanisms for the longitudinal biomarker. For both biological and clinical data, missing longitudinal values are problematic. Thus, a combined method can be more useful to determine if an association exists between longitudinal biomarker values even when some are missing and overall survival which has heterogeneous hazards for different individuals.

Finally, as with most survival methods, informative censoring of the outcome data is problematic regardless of whether or not the hazard functions are heterogeneous. In this dissertation, we did not deal with this issue at all. However, it could possibly be fruitful to model this phenomonom to decrease the bias associated with informative censoring in time to event outcomes. Such methods could have positive implications for both clinical and biological applications.

APPENDIX A

TABLES 1–19
(TABLES 4–1A-4–6B)

Table 1: **Table 4-1A. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ for latent type (1)†**

# of Subjects	# of Time Points	σ_1^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
20	10	1	0.241	0.523	0.754	0.790
20	10	0.8	0.231	0.515	0.732	0.756
20	10	0.2	0.223	0.401	0.428	0.443
25	4	1	0.222	0.570	0.780	0.800
25	4	0.8	0.214	0.554	0.763	0.783
25	4	0.2	0.201	0.506	0.686	0.710
100	20	1	0.205	0.683	0.800	0.830
100	20	0.8	0.193	0.664	0.783	0.820
100	20	0.2	0.181	0.646	0.763	0.810
200	10	1	0.165	0.703	0.810	0.870
200	10	0.8	0.154	0.684	0.794	0.862
200	10	0.2	0.143	0.665	0.773	0.851
500	4	1	0.051	0.726	0.833	0.901
500	4	0.8	0.043	0.715	0.802	0.893
500	4	0.2	0.032	0.703	0.791	0.873
1000	2	1	0.044	0.873	0.901	0.932
1000	2	0.8	0.031	0.864	0.893	0.914
1000	2	0.2	0.021	0.853	0.884	0.903

†Note: $W(t) = U_1$, $U_1 \sim N(0, \sigma_1^2)$

Table 2: **Table 4-1B. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ for latent type (2)†**

# of Subjects	# of Time Points	σ_1^2	σ_2^2	ρ	$\sigma_1^2 + \sigma_2^2\rho$	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
20	10	1	1	0.5	1.5	0.227	0.562	0.776	0.856
20	10	1	1	0.05	1.05	0.225	0.543	0.760	0.803
20	10	0.8	0.2	0.5	0.9	0.218	0.531	0.752	0.784
20	10	0.8	0.2	0.05	0.81	0.216	0.514	0.734	0.691
20	10	0.2	0.8	0.5	0.6	0.203	0.453	0.487	0.503
20	10	0.2	0.8	0.05	0.24	0.201	0.433	0.455	0.472
25	4	1	1	0.5	1.5	0.204	0.614	0.801	0.823
25	4	1	1	0.05	1.05	0.202	0.594	0.784	0.806
25	4	0.8	0.2	0.5	0.9	0.194	0.574	0.773	0.788
25	4	0.8	0.2	0.05	0.81	0.192	0.553	0.752	0.769
25	4	0.2	0.8	0.5	0.6	0.183	0.521	0.697	0.718
25	4	0.2	0.8	0.05	0.24	0.181	0.503	0.672	0.695
100	20	1	1	0.5	1.5	0.179	0.704	0.818	0.855
100	20	1	1	0.05	1.05	0.177	0.688	0.794	0.836
100	20	0.8	0.2	0.5	0.9	0.173	0.686	0.793	0.834
100	20	0.8	0.2	0.05	0.81	0.171	0.665	0.772	0.819
100	20	0.2	0.8	0.5	0.6	0.167	0.647	0.771	0.823
100	20	0.2	0.8	0.05	0.24	0.165	0.630	0.751	0.815

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 3: **Table 4-1B (continued). Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ for latent type (2)**†

# of Subjects	# of Time Points	σ_1^2	σ_2^2	ρ	$\sigma_1^2 + \sigma_2^2\rho$	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
200	10	1	1	0.5	1.5	0.151	0.724	0.828	0.910
200	10	1	1	0.05	1.05	0.149	0.708	0.806	0.892
200	10	0.8	0.2	0.5	0.9	0.147	0.706	0.804	0.893
200	10	0.8	0.2	0.05	0.81	0.145	0.685	0.786	0.876
200	10	0.2	0.8	0.5	0.6	0.143	0.683	0.784	0.874
200	10	0.2	0.8	0.05	0.24	0.141	0.667	0.768	0.856
500	4	1	1	0.5	1.5	0.074	0.766	0.873	0.944
500	4	1	1	0.05	1.05	0.045	0.735	0.831	0.906
500	4	0.8	0.2	0.5	0.9	0.064	0.744	0.853	0.925
500	4	0.8	0.2	0.05	0.81	0.054	0.723	0.814	0.893
500	4	0.2	0.8	0.5	0.6	0.051	0.724	0.834	0.903
500	4	0.2	0.8	0.05	0.24	0.034	0.715	0.805	0.871
1000	2	1	1	0.5	1.5	0.053	0.920	0.944	0.964
1000	2	1	1	0.05	1.05	0.041	0.900	0.925	0.934
1000	2	0.8	0.2	0.5	0.9	0.043	0.913	0.936	0.955
1000	2	0.8	0.2	0.05	0.81	0.034	0.896	0.919	0.926
1000	2	0.2	0.8	0.5	0.6	0.035	0.903	0.925	0.944
1000	2	0.2	0.8	0.05	0.24	0.029	0.883	0.892	0.914

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 4: **Table 4-1C. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ for latent type (3)†**

# of Subjects	# of Time Points	σ_1^2	σ_2^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
20	10	1	1	0.181	0.593	0.791	0.903
20	10	0.8	0.2	0.171	0.565	0.774	0.792
20	10	0.2	0.8	0.162	0.483	0.504	0.525
25	4	1	1	0.173	0.641	0.812	0.842
25	4	0.8	0.2	0.164	0.623	0.781	0.814
25	4	0.2	0.8	0.155	0.568	0.713	0.730
100	20	1	1	0.131	0.727	0.852	0.872
100	20	0.8	0.2	0.123	0.703	0.834	0.866
100	20	0.2	0.8	0.115	0.686	0.816	0.856
200	10	1	1	0.101	0.747	0.852	0.930
200	10	0.8	0.2	0.093	0.723	0.834	0.920
200	10	0.2	0.8	0.086	0.706	0.816	0.911
500	4	1	1	0.051	0.807	0.895	0.975
500	4	0.8	0.2	0.043	0.788	0.876	0.966
500	4	0.2	0.8	0.034	0.765	0.857	0.958
1000	2	1	1	0.042	0.944	0.961	0.990
1000	2	0.8	0.2	0.031	0.935	0.954	0.980
1000	2	0.2	0.8	0.022	0.924	0.943	0.971

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, Corr $(V(t), V(t+s)) = \exp(-|s|)$

Table 5: **Table 4-2A. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ for latent type (1)†**

# of Subjects	# of Time Points	σ_1^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
20	10	1	0.127	0.080	0.121	0.143
20	10	0.8	0.114	0.062	0.093	0.102
20	10	0.2	0.109	0.033	0.062	0.081
25	4	1	0.123	0.086	0.123	0.151
25	4	0.8	0.110	0.065	0.097	0.113
25	4	0.2	0.105	0.037	0.065	0.092
100	20	1	0.104	0.093	0.141	0.183
100	20	0.8	0.098	0.092	0.127	0.162
100	20	0.2	0.092	0.084	0.106	0.141
200	10	1	0.099	0.121	0.161	0.201
200	10	0.8	0.094	0.114	0.144	0.182
200	10	0.2	0.088	0.106	0.122	0.161
500	4	1	0.079	0.132	0.182	0.252
500	4	0.8	0.075	0.123	0.162	0.231
500	4	0.2	0.071	0.111	0.143	0.211
1000	2	1	0.069	0.153	0.213	0.274
1000	2	0.8	0.067	0.142	0.184	0.253
1000	2	0.2	0.063	0.131	0.168	0.231

†Note: $W(t) = U_1$, $U_1 \sim N(0, \sigma_1^2)$,

Table 6: **Table 4-2B. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ for latent type (2)†**

# of Subjects	# of Time Points	σ_1^2	σ_2^2	ρ	$\sigma_1^2 + \sigma_2^2\rho$	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
20	10	1	1	0.5	1.5	0.103	0.132	0.201	0.254
20	10	1	1	0.05	1.05	0.100	0.113	0.182	0.206
20	10	0.8	0.2	0.5	0.9	0.098	0.104	0.144	0.184
20	10	0.8	0.2	0.05	0.81	0.096	0.082	0.123	0.132
20	10	0.2	0.8	0.5	0.6	0.092	0.061	0.084	0.104
20	10	0.2	0.8	0.05	0.24	0.090	0.051	0.062	0.081
25	4	1	1	0.5	1.5	0.097	0.137	0.204	0.258
25	4	1	1	0.05	1.05	0.095	0.118	0.185	0.208
25	4	0.8	0.2	0.5	0.9	0.093	0.109	0.148	0.187
25	4	0.8	0.2	0.05	0.81	0.091	0.088	0.129	0.136
25	4	0.2	0.8	0.5	0.6	0.088	0.069	0.090	0.107
25	4	0.2	0.8	0.05	0.24	0.086	0.056	0.068	0.089
100	20	1	1	0.5	1.5	0.072	0.145	0.222	0.273
100	20	1	1	0.05	1.05	0.070	0.126	0.182	0.256
100	20	0.8	0.2	0.5	0.9	0.067	0.121	0.204	0.251
100	20	0.8	0.2	0.05	0.81	0.065	0.107	0.186	0.232
100	20	0.2	0.8	0.5	0.6	0.062	0.104	0.182	0.237
100	20	0.2	0.8	0.05	0.24	0.060	0.085	0.163	0.211

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 7: **Table 4-2B (continued). Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ for latent type (2)**†

# of Subjects	# of Time Points	σ_1^2	σ_2^2	ρ	$\sigma_1^2 + \sigma_2^2\rho$	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
200	10	1	1	0.5	1.5	0.070	0.162	0.241	0.294
200	10	1	1	0.05	1.05	0.068	0.143	0.226	0.275
200	10	0.8	0.2	0.5	0.9	0.066	0.142	0.225	0.274
200	10	0.8	0.2	0.05	0.81	0.064	0.123	0.203	0.252
200	10	0.2	0.8	0.5	0.6	0.062	0.121	0.201	0.251
200	10	0.2	0.8	0.05	0.24	0.060	0.103	0.182	0.234
500	4	1	1	0.5	1.5	0.060	0.194	0.275	0.332
500	4	1	1	0.05	1.05	0.058	0.175	0.255	0.304
500	4	0.8	0.2	0.5	0.9	0.057	0.186	0.243	0.303
500	4	0.8	0.2	0.05	0.81	0.055	0.156	0.236	0.286
500	4	0.2	0.8	0.5	0.6	0.054	0.165	0.227	0.284
500	4	0.2	0.8	0.05	0.24	0.052	0.135	0.206	0.263
1000	2	1	1	0.5	1.5	0.056	0.213	0.293	0.351
1000	2	1	1	0.05	1.05	0.054	0.195	0.274	0.324
1000	2	0.8	0.2	0.5	0.9	0.053	0.204	0.264	0.332
1000	2	0.8	0.2	0.05	0.81	0.051	0.176	0.249	0.302
1000	2	0.2	0.8	0.5	0.6	0.050	0.185	0.246	0.303
1000	2	0.2	0.8	0.05	0.24	0.048	0.153	0.221	0.285

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 8: **Table 4-2C. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ for latent type (3)†**

# of Subjects	# of Time Points	σ_1^2	σ_2^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
20	10	1	1	0.088	0.162	0.226	0.278
20	10	0.8	0.2	0.084	0.136	0.164	0.199
20	10	0.2	0.8	0.080	0.117	0.127	0.136
25	4	1	1	0.079	0.167	0.230	0.282
25	4	0.8	0.2	0.075	0.140	0.169	0.203
25	4	0.2	0.8	0.071	0.131	0.131	0.141
100	20	1	1	0.066	0.193	0.255	0.297
100	20	0.8	0.2	0.063	0.176	0.237	0.279
100	20	0.2	0.8	0.060	0.153	0.216	0.254
200	10	1	1	0.058	0.214	0.276	0.318
200	10	0.8	0.2	0.055	0.197	0.255	0.299
200	10	0.2	0.8	0.052	0.177	0.237	0.276
500	4	1	1	0.052	0.238	0.297	0.354
500	4	0.8	0.2	0.049	0.227	0.278	0.335
500	4	0.2	0.8	0.046	0.209	0.257	0.316
1000	2	1	1	0.048	0.253	0.312	0.383
1000	2	0.8	0.2	0.045	0.242	0.297	0.352
1000	2	0.2	0.8	0.042	0.220	0.275	0.336

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 9: **Table 4-3. Measure of marker effectiveness $R_{M_1}(\tau_1, \tau_2)$ for $W(t)$ model for latent type (3)† under the different correlation between survival time and biomarker {Fixed Point Method}**

# of Subjects	# of Time Points	Percentage of Censor	σ_1^2	σ_2^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
20	10	0	1	1	0.01 ± 0.047	0.03 ± 0.049	0.10 ± 0.050	0.30 ± 0.052
25	4	0	1	1	0.01 ± 0.049	0.02 ± 0.050	0.08 ± 0.049	0.25 ± 0.052
100	20	0	1	1	0.10 ± 0.049	0.15 ± 0.049	0.30 ± 0.049	0.65 ± 0.050
200	10	0	1	1	0.08 ± 0.050	0.13 ± 0.050	0.27 ± 0.049	0.60 ± 0.051
500	4	0	1	1	0.06 ± 0.051	0.11 ± 0.048	0.24 ± 0.050	0.55 ± 0.051
1000	2	0	1	1	0.04 ± 0.049	0.09 ± 0.048	0.21 ± 0.049	0.50 ± 0.051
20	10	10	1	1	0.009 ± 0.047	0.015 ± 0.049	0.08 ± 0.050	0.25 ± 0.052
25	4	10	1	1	0.009 ± 0.049	0.01 ± 0.050	0.06 ± 0.049	0.20 ± 0.052
100	20	10	1	1	0.08 ± 0.049	0.13 ± 0.049	0.25 ± 0.049	0.55 ± 0.050
200	10	10	1	1	0.06 ± 0.050	0.11 ± 0.050	0.22 ± 0.049	0.50 ± 0.051
500	4	10	1	1	0.04 ± 0.051	0.09 ± 0.048	0.19 ± 0.050	0.45 ± 0.051
1000	2	10	1	1	0.02 ± 0.049	0.07 ± 0.048	0.16 ± 0.049	0.40 ± 0.051
20	10	25	1	1	0.007 ± 0.047	0.01 ± 0.049	0.06 ± 0.050	0.20 ± 0.052
25	4	25	1	1	0.007 ± 0.049	0.008 ± 0.050	0.04 ± 0.049	0.15 ± 0.052
100	20	25	1	1	0.06 ± 0.049	0.11 ± 0.049	0.21 ± 0.049	0.45 ± 0.050
200	10	25	1	1	0.04 ± 0.050	0.09 ± 0.050	0.19 ± 0.049	0.40 ± 0.051
500	4	25	1	1	0.02 ± 0.051	0.07 ± 0.048	0.17 ± 0.050	0.35 ± 0.051
1000	2	25	1	1	0.01 ± 0.049	0.05 ± 0.048	0.15 ± 0.049	0.30 ± 0.051

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 10: **Table 4-3 (continued). Measure of marker effectiveness $R_{M_1}(\tau_1, \tau_2)$ for $W(t)$ model for latent type (3)† under the different correlation between survival time and biomarker {Fixed Point Method}**

# of Subjects	# of Time Points	Percentage of Censor	σ_1^2	σ_2^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
20	10	50	1	1	0.005 ± 0.047	0.008 ± 0.049	0.04 ± 0.050	0.15 ± 0.052
25	4	50	1	1	0.005 ± 0.049	0.006 ± 0.050	0.02 ± 0.049	0.10 ± 0.052
100	20	50	1	1	0.04 ± 0.049	0.09 ± 0.049	0.17 ± 0.049	0.35 ± 0.050
200	10	50	1	1	0.02 ± 0.050	0.07 ± 0.050	0.15 ± 0.049	0.30 ± 0.051
500	4	50	1	1	0.01 ± 0.051	0.05 ± 0.048	0.13 ± 0.050	0.25 ± 0.051
1000	2	50	1	1	0.008 ± 0.049	0.03 ± 0.048	0.11 ± 0.049	0.20 ± 0.051

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 11: **Table 4-4. Measure of marker effectiveness $R_{D_1}(\tau_1, \tau_2)$ for $W(t)$ model for latent type (3)† under the different correlation between survival time and biomarker {Interval Measure Method}**

# of Subjects	# of Time Points	Percentage of Censor	σ_1^2	σ_2^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
20	10	0	1	1	0.01 ± 0.047	0.03 ± 0.049	0.10 ± 0.050	0.30 ± 0.052
25	4	0	1	1	0.01 ± 0.049	0.02 ± 0.050	0.08 ± 0.049	0.25 ± 0.052
100	20	0	1	1	0.10 ± 0.049	0.15 ± 0.049	0.30 ± 0.049	0.65 ± 0.050
200	10	0	1	1	0.08 ± 0.050	0.13 ± 0.050	0.27 ± 0.049	0.60 ± 0.051
500	4	0	1	1	0.06 ± 0.051	0.11 ± 0.048	0.24 ± 0.050	0.55 ± 0.051
1000	2	0	1	1	0.04 ± 0.049	0.09 ± 0.048	0.21 ± 0.049	0.50 ± 0.051
20	10	10	1	1	0.0095 ± 0.047	0.02 ± 0.049	0.09 ± 0.050	0.28 ± 0.052
25	4	10	1	1	0.0095 ± 0.049	0.015 ± 0.050	0.07 ± 0.049	0.23 ± 0.052
100	20	10	1	1	0.09 ± 0.049	0.14 ± 0.049	0.27 ± 0.049	0.60 ± 0.050
200	10	10	1	1	0.07 ± 0.050	0.12 ± 0.050	0.24 ± 0.049	0.55 ± 0.051
500	4	10	1	1	0.05 ± 0.051	0.10 ± 0.048	0.21 ± 0.050	0.50 ± 0.051
1000	2	10	1	1	0.03 ± 0.049	0.08 ± 0.048	0.18 ± 0.049	0.45 ± 0.051
20	10	25	1	1	0.008 ± 0.047	0.015 ± 0.049	0.08 ± 0.050	0.26 ± 0.052
25	4	25	1	1	0.008 ± 0.049	0.01 ± 0.050	0.06 ± 0.049	0.21 ± 0.052
100	20	25	1	1	0.07 ± 0.049	0.13 ± 0.049	0.24 ± 0.049	0.55 ± 0.050
200	10	25	1	1	0.05 ± 0.050	0.11 ± 0.050	0.22 ± 0.049	0.50 ± 0.051
500	4	25	1	1	0.03 ± 0.051	0.09 ± 0.048	0.20 ± 0.050	0.45 ± 0.051
1000	2	25	1	1	0.02 ± 0.049	0.07 ± 0.048	0.18 ± 0.049	0.40 ± 0.051

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 12: **Table 4-4 (continued). Measure of marker effectiveness $R_{D_1}(\tau_1, \tau_2)$ for $W(t)$ model for latent type (3)† under the different correlation between survival time and biomarker {Interval Measure Method}**

# of Subjects	# of Time Points	Percentage of Censor	σ_1^2	σ_2^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
20	10	50	1	1	0.007 ± 0.047	0.01 ± 0.049	0.05 ± 0.050	0.24 ± 0.052
25	4	50	1	1	0.007 ± 0.049	0.008 ± 0.050	0.05 ± 0.049	0.19 ± 0.052
100	20	50	1	1	0.05 ± 0.049	0.12 ± 0.049	0.21 ± 0.049	0.50 ± 0.050
200	10	50	1	1	0.03 ± 0.050	0.10 ± 0.050	0.19 ± 0.049	0.45 ± 0.051
500	4	50	1	1	0.02 ± 0.051	0.08 ± 0.048	0.17 ± 0.050	0.40 ± 0.051
1000	2	50	1	1	0.01 ± 0.049	0.06 ± 0.048	0.15 ± 0.049	0.35 ± 0.051

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 13: **Table 4-5A. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (1)**†

# of Subjects	# of Time Points	σ_1^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
100	20	1	0.216	0.674	0.791	0.822
100	20	0.8	0.202	0.653	0.772	0.813
100	20	0.2	0.193	0.635	0.754	0.801
200	10	1	0.174	0.692	0.803	0.864
200	10	0.8	0.162	0.677	0.782	0.851
200	10	0.2	0.154	0.654	0.761	0.843
500	4	1	0.060	0.718	0.824	0.892
500	4	0.8	0.052	0.703	0.793	0.881
500	4	0.2	0.044	0.691	0.782	0.865
1000	2	1	0.052	0.865	0.890	0.924
1000	2	0.8	0.042	0.852	0.884	0.892
1000	2	0.2	0.030	0.845	0.872	0.881

†Note: $W(t) = U_1$, $U_1 \sim N(0, \sigma_1^2)$

Table 14: **Table 4-5B. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (2)**†

# of Subjects	# of Time Points	σ_1^2	σ_2^2	ρ	$\sigma_1^2 + \sigma_2^2\rho$	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
100	20	1	1	0.5	1.5	0.187	0.695	0.806	0.846
100	20	1	1	0.05	1.05	0.186	0.676	0.782	0.823
100	20	0.8	0.2	0.5	0.9	0.184	0.674	0.781	0.825
100	20	0.8	0.2	0.05	0.81	0.183	0.657	0.764	0.807
100	20	0.2	0.8	0.5	0.6	0.175	0.665	0.762	0.811
100	20	0.2	0.8	0.05	0.24	0.172	0.622	0.743	0.803
200	10	1	1	0.5	1.5	0.163	0.712	0.817	0.901
200	10	1	1	0.05	1.05	0.158	0.695	0.795	0.881
200	10	0.8	0.2	0.5	0.9	0.155	0.694	0.792	0.880
200	10	0.8	0.2	0.05	0.81	0.153	0.673	0.775	0.864
200	10	0.2	0.8	0.5	0.6	0.151	0.671	0.772	0.862
200	10	0.2	0.8	0.05	0.24	0.150	0.655	0.755	0.843

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 15: **Table 4-5B (continued). Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (2)†**

# of Subjects	# of Time Points	σ_1^2	σ_2^2	ρ	$\sigma_1^2 + \sigma_2^2\rho$	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
500	4	1	1	0.5	1.5	0.082	0.754	0.865	0.932
500	4	1	1	0.05	1.05	0.053	0.722	0.822	0.894
500	4	0.8	0.2	0.5	0.9	0.072	0.736	0.841	0.913
500	4	0.8	0.2	0.05	0.81	0.062	0.711	0.802	0.881
500	4	0.2	0.8	0.5	0.6	0.060	0.712	0.822	0.892
500	4	0.2	0.8	0.05	0.24	0.042	0.702	0.793	0.863
1000	2	1	1	0.5	1.5	0.061	0.911	0.931	0.952
1000	2	1	1	0.05	1.05	0.050	0.892	0.913	0.923
1000	2	0.8	0.2	0.5	0.9	0.051	0.901	0.922	0.942
1000	2	0.8	0.2	0.05	0.81	0.042	0.887	0.907	0.917
1000	2	0.2	0.8	0.5	0.6	0.046	0.892	0.913	0.932
1000	2	0.2	0.8	0.05	0.24	0.037	0.871	0.881	0.903

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 16: **Table 4-5C. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (3)**†

# of Subjects	# of Time Points	σ_1^2	σ_2^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
100	20	1	1	0.140	0.715	0.844	0.864
100	20	0.8	0.2	0.131	0.691	0.822	0.854
100	20	0.2	0.8	0.123	0.673	0.804	0.843
200	10	1	1	0.110	0.735	0.843	0.922
200	10	0.8	0.2	0.102	0.711	0.822	0.911
200	10	0.2	0.8	0.093	0.693	0.804	0.903
500	4	1	1	0.063	0.795	0.883	0.963
500	4	0.8	0.2	0.051	0.775	0.863	0.954
500	4	0.2	0.8	0.042	0.753	0.844	0.946
1000	2	1	1	0.055	0.932	0.952	0.982
1000	2	0.8	0.2	0.040	0.923	0.942	0.973
1000	2	0.2	0.8	0.031	0.912	0.931	0.960

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, Corr $(V(t), V(t+s)) = \exp(-|s|)$

Table 17: **Table 4-6A. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (1)†**

# of Subjects	# of Time Points	σ_1^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
100	20	1	0.115	0.081	0.123	0.161
100	20	0.8	0.105	0.080	0.104	0.141
100	20	0.2	0.101	0.072	0.083	0.120
200	10	1	0.107	0.113	0.153	0.193
200	10	0.8	0.103	0.102	0.131	0.170
200	10	0.2	0.096	0.093	0.114	0.153
500	4	1	0.087	0.124	0.170	0.241
500	4	0.8	0.083	0.111	0.151	0.223
500	4	0.2	0.080	0.104	0.134	0.204
1000	2	1	0.077	0.141	0.201	0.262
1000	2	0.8	0.075	0.130	0.172	0.241
1000	2	0.2	0.071	0.123	0.155	0.223

†Note: $W(t) = U_1$, $U_1 \sim N(0, \sigma_1^2)$,

Table 18: **Table 4-6B. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (2)**†

# of Subjects	# of Time Points	σ_1^2	σ_2^2	ρ	$\sigma_1^2 + \sigma_2^2\rho$	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
100	20	1	1	0.5	1.5	0.084	0.135	0.213	0.265
100	20	1	1	0.05	1.05	0.082	0.116	0.170	0.243
100	20	0.8	0.2	0.5	0.9	0.075	0.111	0.192	0.243
100	20	0.8	0.2	0.05	0.81	0.073	0.097	0.173	0.223
100	20	0.2	0.8	0.5	0.6	0.071	0.094	0.171	0.223
100	20	0.2	0.8	0.05	0.24	0.069	0.075	0.152	0.203
200	10	1	1	0.5	1.5	0.080	0.153	0.232	0.282
200	10	1	1	0.05	1.05	0.078	0.131	0.214	0.266
200	10	0.8	0.2	0.5	0.9	0.076	0.130	0.212	0.262
200	10	0.8	0.2	0.05	0.81	0.074	0.111	0.194	0.243
200	10	0.2	0.8	0.5	0.6	0.072	0.110	0.192	0.242
200	10	0.2	0.8	0.05	0.24	0.070	0.092	0.170	0.223

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 19: **Table 4-6B (continued). Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (2)†**

# of Subjects	# of Time Points	σ_1^2	σ_2^2	ρ	$\sigma_1^2 + \sigma_2^2\rho$	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
500	4	1	1	0.5	1.5	0.072	0.186	0.267	0.324
500	4	1	1	0.05	1.05	0.066	0.167	0.243	0.293
500	4	0.8	0.2	0.5	0.9	0.064	0.178	0.235	0.292
500	4	0.8	0.2	0.05	0.81	0.063	0.143	0.224	0.273
500	4	0.2	0.8	0.5	0.6	0.062	0.154	0.215	0.272
500	4	0.2	0.8	0.05	0.24	0.060	0.123	0.194	0.251
1000	2	1	1	0.5	1.5	0.063	0.201	0.281	0.342
1000	2	1	1	0.05	1.05	0.062	0.183	0.262	0.312
1000	2	0.8	0.2	0.5	0.9	0.062	0.192	0.257	0.320
1000	2	0.8	0.2	0.05	0.81	0.060	0.164	0.236	0.291
1000	2	0.2	0.8	0.5	0.6	0.059	0.173	0.234	0.291
1000	2	0.2	0.8	0.05	0.24	0.057	0.142	0.210	0.273

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

APPENDIX B

TABLES 20–38
(TABLES 4–6C-4–9(C2))

Table 20: **Table 4-6C. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (3)**†

# of Subjects	# of Time Points	σ_1^2	σ_2^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
100	20	1	1	0.073	0.181	0.243	0.285
100	20	0.8	0.2	0.072	0.164	0.225	0.266
100	20	0.2	0.8	0.071	0.142	0.204	0.242
200	10	1	1	0.066	0.203	0.263	0.305
200	10	0.8	0.2	0.063	0.185	0.243	0.286
200	10	0.2	0.8	0.061	0.164	0.225	0.264
500	4	1	1	0.060	0.226	0.285	0.342
500	4	0.8	0.2	0.057	0.215	0.266	0.323
500	4	0.2	0.8	0.054	0.198	0.245	0.304
1000	2	1	1	0.056	0.242	0.300	0.371
1000	2	0.8	0.2	0.054	0.231	0.285	0.343
1000	2	0.2	0.8	0.050	0.211	0.264	0.324

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 21: **Table 4-7A. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (1)†**

# of Subjects	# of Time Points	σ_1^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
100	20	1	0.317	0.374	0.396	0.423
100	20	0.8	0.304	0.356	0.371	0.414
100	20	0.2	0.292	0.337	0.352	0.402
200	10	1	0.276	0.395	0.403	0.464
200	10	0.8	0.262	0.373	0.382	0.451
200	10	0.2	0.251	0.353	0.361	0.444
500	4	1	0.163	0.417	0.424	0.493
500	4	0.8	0.152	0.402	0.393	0.480
500	4	0.2	0.145	0.394	0.380	0.464
1000	2	1	0.152	0.564	0.593	0.625
1000	2	0.8	0.145	0.552	0.584	0.591
1000	2	0.2	0.133	0.545	0.575	0.585

†Note: $W(t) = U_1$, $U_1 \sim N(0, \sigma_1^2)$

Table 22: **Table 4-7B. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (2)**†

# of Subjects	# of Time Points	σ_1^2	σ_2^2	ρ	$\sigma_1^2 + \sigma_2^2\rho$	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
100	20	1	1	0.5	1.5	0.288	0.497	0.607	0.647
100	20	1	1	0.05	1.05	0.285	0.474	0.582	0.624
100	20	0.8	0.2	0.5	0.9	0.284	0.471	0.581	0.622
100	20	0.8	0.2	0.05	0.81	0.280	0.456	0.564	0.607
100	20	0.2	0.8	0.5	0.6	0.278	0.463	0.562	0.611
100	20	0.2	0.8	0.05	0.24	0.272	0.422	0.540	0.603
200	10	1	1	0.5	1.5	0.263	0.514	0.612	0.705
200	10	1	1	0.05	1.05	0.257	0.496	0.594	0.681
200	10	0.8	0.2	0.5	0.9	0.256	0.493	0.596	0.680
200	10	0.8	0.2	0.05	0.81	0.254	0.472	0.573	0.663
200	10	0.2	0.8	0.5	0.6	0.252	0.471	0.572	0.662
200	10	0.2	0.8	0.05	0.24	0.250	0.453	0.556	0.644

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 23: **Table 4-7B (continued). Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (2)†**

# of Subjects	# of Time Points	σ_1^2	σ_2^2	ρ	$\sigma_1^2 + \sigma_2^2\rho$	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
500	4	1	1	0.5	1.5	0.182	0.558	0.662	0.731
500	4	1	1	0.05	1.05	0.154	0.526	0.623	0.697
500	4	0.8	0.2	0.5	0.9	0.175	0.533	0.642	0.712
500	4	0.8	0.2	0.05	0.81	0.163	0.512	0.602	0.681
500	4	0.2	0.8	0.5	0.6	0.160	0.510	0.623	0.696
500	4	0.2	0.8	0.05	0.24	0.147	0.502	0.592	0.660
1000	2	1	1	0.5	1.5	0.164	0.712	0.731	0.752
1000	2	1	1	0.05	1.05	0.152	0.693	0.716	0.727
1000	2	0.8	0.2	0.5	0.9	0.151	0.701	0.723	0.743
1000	2	0.8	0.2	0.05	0.81	0.142	0.687	0.705	0.712
1000	2	0.2	0.8	0.5	0.6	0.143	0.692	0.717	0.731
1000	2	0.2	0.8	0.05	0.24	0.137	0.671	0.683	0.702

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 24: **Table 4-7C. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (3)**†

# of Subjects	# of Time Points	σ_1^2	σ_2^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
100	20	1	1	0.242	0.515	0.645	0.665
100	20	0.8	0.2	0.232	0.494	0.622	0.653
100	20	0.2	0.8	0.227	0.473	0.603	0.647
200	10	1	1	0.213	0.535	0.644	0.722
200	10	0.8	0.2	0.201	0.514	0.621	0.713
200	10	0.2	0.8	0.197	0.493	0.602	0.700
500	4	1	1	0.164	0.594	0.687	0.962
500	4	0.8	0.2	0.152	0.575	0.663	0.757
500	4	0.2	0.8	0.141	0.552	0.644	0.744
1000	2	1	1	0.154	0.737	0.753	0.982
1000	2	0.8	0.2	0.142	0.723	0.742	0.774
1000	2	0.2	0.8	0.135	0.716	0.735	0.760

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, Corr $(V(t), V(t+s)) = \exp(-|s|)$

Table 25: **Table 4-8A. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (1)†**

# of Subjects	# of Time Points	σ_1^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
100	20	1	0.216	0.086	0.125	0.167
100	20	0.8	0.209	0.083	0.103	0.144
100	20	0.2	0.204	0.071	0.089	0.122
200	10	1	0.207	0.114	0.154	0.194
200	10	0.8	0.202	0.102	0.138	0.171
200	10	0.2	0.196	0.097	0.110	0.150
500	4	1	0.187	0.124	0.173	0.244
500	4	0.8	0.183	0.111	0.151	0.226
500	4	0.2	0.180	0.103	0.135	0.203
1000	2	1	0.177	0.146	0.202	0.262
1000	2	0.8	0.174	0.133	0.176	0.244
1000	2	0.2	0.171	0.122	0.154	0.222

†Note: $W(t) = U_1$, $U_1 \sim N(0, \sigma_1^2)$,

Table 26: **Table 4-8B. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (2)**†

# of Subjects	# of Time Points	σ_1^2	σ_2^2	ρ	$\sigma_1^2 + \sigma_2^2\rho$	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
100	20	1	1	0.5	1.5	0.184	0.137	0.155	0.165
100	20	1	1	0.05	1.05	0.182	0.114	0.123	0.144
100	20	0.8	0.2	0.5	0.9	0.173	0.112	0.147	0.140
100	20	0.8	0.2	0.05	0.81	0.172	0.095	0.123	0.125
100	20	0.2	0.8	0.5	0.6	0.171	0.092	0.125	0.121
100	20	0.2	0.8	0.05	0.24	0.169	0.077	0.107	0.105
200	10	1	1	0.5	1.5	0.183	0.152	0.145	0.187
200	10	1	1	0.05	1.05	0.179	0.134	0.152	0.163
200	10	0.8	0.2	0.5	0.9	0.175	0.132	0.153	0.165
200	10	0.8	0.2	0.05	0.81	0.171	0.112	0.144	0.143
200	10	0.2	0.8	0.5	0.6	0.170	0.111	0.143	0.140
200	10	0.2	0.8	0.05	0.24	0.169	0.093	0.131	0.127

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 27: **Table 4-8B (continued). Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (2)†**

# of Subjects	# of Time Points	σ_1^2	σ_2^2	ρ	$\sigma_1^2 + \sigma_2^2\rho$	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
500	4	1	1	0.5	1.5	0.171	0.186	0.216	0.227
500	4	1	1	0.05	1.05	0.167	0.167	0.183	0.193
500	4	0.8	0.2	0.5	0.9	0.166	0.173	0.164	0.192
500	4	0.8	0.2	0.05	0.81	0.164	0.144	0.153	0.173
500	4	0.2	0.8	0.5	0.6	0.165	0.157	0.158	0.171
500	4	0.2	0.8	0.05	0.24	0.162	0.122	0.144	0.157
1000	2	1	1	0.5	1.5	0.166	0.205	0.206	0.246
1000	2	1	1	0.05	1.05	0.164	0.187	0.192	0.223
1000	2	0.8	0.2	0.5	0.9	0.163	0.192	0.188	0.220
1000	2	0.8	0.2	0.05	0.81	0.161	0.167	0.167	0.197
1000	2	0.2	0.8	0.5	0.6	0.160	0.173	0.178	0.191
1000	2	0.2	0.8	0.05	0.24	0.158	0.141	0.154	0.170

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 28: **Table 4-8C. Power of score test for the $W(t)$ model under $H_0: \gamma = 0$ given 50 % longitudinal biomarkers missing for latent type (3)**†

# of Subjects	# of Time Points	σ_1^2	σ_2^2	$\gamma = 0.00$	$\gamma = 0.10$	$\gamma = 0.25$	$\gamma = 0.85$
100	20	1	1	0.174	0.187	0.206	0.229
100	20	0.8	0.2	0.173	0.166	0.184	0.206
100	20	0.2	0.8	0.169	0.143	0.162	0.181
200	10	1	1	0.165	0.207	0.228	0.244
200	10	0.8	0.2	0.163	0.187	0.204	0.226
200	10	0.2	0.8	0.160	0.167	0.182	0.204
500	4	1	1	0.158	0.229	0.245	0.261
500	4	0.8	0.2	0.157	0.214	0.234	0.254
500	4	0.2	0.8	0.154	0.195	0.216	0.239
1000	2	1	1	0.151	0.247	0.264	0.287
1000	2	0.8	0.2	0.148	0.230	0.253	0.270
1000	2	0.2	0.8	0.145	0.214	0.237	0.252

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 29: **Table 4-9(A1). Liver cirrhosis trial results for the $W(t)$ model for latent type (1)† under $H_0: \ \gamma = 0$**

	Est	SE
Treatment	0.586	0.7125
Random effect from prothrombin	-0.162	0.0182

frailty effect: Chisq = 17633.87, DF = 427

†Note: $W(t) = U_1$, $U_1 \sim N(0, \sigma_1^2)$

Table 30: **Table 4-9(A2). Score test for the $W(t)$ model for latent type (1)† under $H_0: \ \gamma = 0$**

Chisq	DF	p
181	29	< 0.001

†Note: $W(t) = U_1$, $U_1 \sim N(0, \sigma_1^2)$

Table 31: **Table 4-9(A3). Liver cirrhosis trial results for the $W(t)$ model for latent type (1)† under $H_0: \ \gamma = 0$ without frailty**

	Est	SE
Treatment	0.0767	0.11742
Random effect from prothrombin	-0.0438	0.00346

†Note: $W(t) = U_1$, $U_1 \sim N(0, \sigma_1^2)$

Table 32: **Table 4-9(A4). Score test for the $W(t)$ model for latent type (1)† under $H_0: \ \gamma = 0$ without frailty**

Chisq	DF	p
166.49	1	< 0.001

†Note: $W(t) = U_1$, $U_1 \sim N(0, \sigma_1^2)$

Table 33: **Table 4-9(B1). Score test for the $W(t)$ for latent type (2)† model under** $H_0: \ \gamma = 0$

	Est	SE
Treatment	-0.319	0.5504
Random effect from prothrombin	-0.423	0.0827

frailty effect: Chisq = 14652.33, DF = 436

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 34: **Table 4-9(B2). Score test for the $W(t)$ model for latent type (2)† under** $H_0: \ \gamma = 0$

Chisq	DF	p
272	19	< 0.001

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 35: **Table 4-9(B3). Liver cirrhosis trial results for the $W(t)$ model for latent type (2)† under $H_0: \ \gamma = 0$ without frailty**

	Est	SE
Treatment	0.0832	0.1179
Random effect from prothrombin	-0.5350	0.0311

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 36: **Table 4-9(B4). Score test for the $W(t)$ model for latent type (2)† under $H_0: \ \gamma = 0$ without frailty**

Chisq	DF	p
374.49	1	< 0.001

†Note: $W(t) = U_1$, $U_1 \sim N(0, \sigma_1^2)$

Table 37: **Table 4-9(C1). Score test for the $W(t)$ model for latent type (3)† under** $H_0: \ \gamma = 0$

	Est	SE
Treatment	-0.319	0.5504
Random effect from prothrombin	-0.268	0.0523

frailty effect: Chisq = 14652.18, DF = 436

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 38: **Table 4-9(C2). Score test for the $W(t)$ model for latent type (3)† under** $H_0: \ \gamma = 0$

Chisq	DF	p
272	19	< 0.001

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

APPENDIX C

TABLES 39–52
(TABLES 4–9(C3)-4–11(C2))

Table 39: **Table 4-9(C3). Liver cirrhosis trial results for the $W(t)$ model for latent type (3)† under $H_0: \ \gamma = 0$ without frailty**

	Est	SE
Treatment	0.0832	0.1179
Random effect from prothrombin	-0.3388	0.0197

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 40: **Table 4-9(C4). Score test for the $W(t)$ model for latent type (3)† under $H_0: \ \gamma = 0$ without frailty**

Chisq	DF	p
374.49	1	< 0.001

†Note: $W(t) = U_1$, $U_1 \sim N(0, \sigma_1^2)$

Table 41: **Table 4-10(A1). Measure of marker effectiveness for $W(t)$ model for latent type (1)† under the different correlation between survival time and biomarker {Fixed Point Method}**

Time	M_{Y_1}	M_1	R_{M_1}
[0, 1)	0.559	0.738	0.243
[3, 4)	0.817	0.817	0.000

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 42: **Table 4-10(A2). Measure of marker effectiveness for $W(t)$ model for latent type (1)† under the different correlation between survival time and biomarker without frailty{Fixed Point Method}**

Time	M_{Y_1}	M_1	R_{M_1}
[0, 1)	0.242	0.252	0.038
[3, 4)	0.583	0.583	0.000

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 43: **Table 4-10(B1). Measure of marker effectiveness for $W(t)$ model for latent type (2)† under the different correlation between survival time and biomarker {Fixed Point Method}**

Time	M_{Y_1}	M_1	R_{M_1}
[0, 1)	0.510	0.738	0.310
[3, 4)	0.817	0.817	0.000

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 44: **Table 4-10(B2). Measure of marker effectiveness for $W(t)$ model for latent type (2)† under the different correlation between survival time and biomarker {Fixed Point Method}**

Time	M_{Y_1}	M_1	R_{M_1}
[0, 1)	0.203	0.252	0.294
[3, 4)	0.583	0.583	0.000

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 45: **Table 4-10(C1). Measure of marker effectiveness for $W(t)$ model for latent type (3)† under the different correlation between survival time and biomarker {Fixed Point Method}**

Time	M_{Y_1}	M_1	R_{M_1}
[0, 1)	0.510	0.738	0.310
[3, 4)	0.817	0.817	0.000

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 46: **Table 4-10(C2). Measure of marker effectiveness for $W(t)$ model for latent type (3)† under the different correlation between survival time and biomarker {Fixed Point Method}**

Time	M_{Y_1}	M_1	R_{M_1}
[0, 1)	0.203	0.252	0.294
[3, 4)	0.583	0.583	0.000

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 47: **Table 4-11(A1). Measure of marker effectiveness for $W(t)$ model for latent type (1)† under the different correlation between survival time and biomarker {Interval Measure Method}**

Time	D_{Y_1}	D_1	R_{D_1}
[0, 1)	0.134	0.511	0.738
[3, 4)	0.817	0.817	0.000

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 48: **Table 4-11(A2). Measure of marker effectiveness for $W(t)$ model for latent type (1)† under the different correlation between survival time and biomarker {Interval Measure Method}**

Time	D_{Y_1}	D_1	R_{D_1}
[0, 1)	0.016	0.052	0.690
[3, 4)	0.583	0.583	0.000

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 49: **Table 4-11(B1). Measure of marker effectiveness for $W(t)$ model for latent type (2)† under the different correlation between survival time and biomarker {Interval Measure Method}**

Time	D_{Y_1}	D_1	R_{D_1}
[0, 1)	0.119	0.511	0.768
[3, 4)	0.817	0.817	0.000

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 50: **Table 4-11(B2). Measure of marker effectiveness for $W(t)$ model for latent type (2)† under the different correlation between survival time and biomarker {Interval Measure Method}**

Time	D_{Y_1}	D_1	R_{D_1}
[0, 1)	0.014	0.052	0.731
[3, 4)	0.583	0.583	0.000

†Note: $W(t) = U_1 + U_2 \times t$, $U_1 \sim N(0, \sigma_1^2)$, $U_2 \sim N(0, \sigma_2^2)$, $Corr(U_1, U_2) = \rho$

Table 51: **Table 4-11(C1). Measure of marker effectiveness for $W(t)$ model for latent type (3)† under the different correlation between survival time and biomarker {Interval Measure Method}**

Time	D_{Y_1}	D_1	R_{D_1}
[0, 1)	0.119	0.511	0.768
[3, 4)	0.817	0.817	0.000

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

Table 52: **Table 4-11(C2). Measure of marker effectiveness for $W(t)$ model for latent type (3)† under the different correlation between survival time and biomarker {Interval Measure Method}**

Time	D_{Y_1}	D_1	R_{D_1}
[0, 1)	0.014	0.052	0.731
[3, 4)	0.583	0.583	0.000

†Note: $W(t) = U_1 + V(t)$, $U_1 \sim N(0, \sigma_1^2)$, $V(t) \sim N(0, \sigma_1^2)$, $Corr(V(t), V(t+s)) = exp(-|s|)$

BIBLIOGRAPHY

1. Paterson A., Szafran O., Hanson J., and Cyr M., Lees A. (1985). Response to treatment and its influence on survival in metastatic breast cancer. Am J Clin Oncol (CCT) 8, 283-292.

2. Gail M. H. (1981). Evaluating serial cancer marker studies in patients at risk of recurrent disease. Biometrics 37, 67-78.

3. Tsiatis A. A., DeGruttola V., and Wulfsohn M.S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. JASA 90(429), 27-37.

4. Weiss G. B., Bunce H., and Hokanson J. A. (1983). Comparing survival of responders and nonresponders after treatment: A potential source of confusion in interpreting cancer clinical trial. Controlled Clinical trial 4, 43-52.

5. Wittes J., Lakatos E., and Probstfield J. (1989). Surrogate endpoints in clinical trials: Cardiovascular disease. Statistics in Medicine 8, 415-425.

6. Freedman S., Graubard B. I., and Schatzkin A. (1992). Statistical validation of intermediate endpoints for chronic disease. Statistics in Medicine 11, 167-178.

7. Pepe M. S. (1992). Inference using surrogate outcome data and a validation sample. Biometrika 79(2), 355-65.

8. Buyse M., Molenberghs G., Burzykowski T., Renard D., and Gey H (2000). The validation of surrogate endpoints in meta-analysis of randomized experiments. Biostatistics 1(1), 49-67.

9. Prentice R. L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. Statistics in Medicine 8, 431-440.

10. Buyse M., and Molenberghs G. (1998). Criteria for the validation of surrogate endpoints in randomized experiments. Biometrics 54(3), 1014-1029.

11. Gail M. H., Pferffer R., Van Houwelingen H. C., and Carroll R. J. (2000). On meta-analytic assessment of surrogate outcomes. Biostatistics 1(3), 231-246.

12. Bruzzi P. (2002). Objective response to treatment as a potential surrogate marker of survival in breast cancer. Ann. Sci. 963 144-147.

13. Hu P., Tsiatis A. A., and Davidian M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. Biometrics 54(12), 1407-1419.

14. Tsiatis A. A. and Davidian M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. Biometrika 88(2), 447-458.

15. Liu K., Mazumdar S., Stone R. A., Dew M. A., Houck P. R., Reynold III C. F. (2001). Accounting for covariate measurement error in a Cox model analysis of recurrence of depression. Journal of Psychiatric Research 35, 177-185.

16. Wulfsohn M. S., and Tsiatis A. A. (1997). A joint model for survival and longitudinal data measured with error. Biometrics 53(1), 330-339.

17. Henderson R., Diggle P., and Dobson A. (2000). Joint modelling of longitudinal measurements and event time data. Biostatistics 1(4), 465-480.

18. Huang Y., and Louis T. A. (1998). Nonparametric estimation of the joint distribution of survival time and mark variables. Biometrika 85(4), 785-798.

19. Lin H., Turnbull B. W., McCulloch C. H., and Slate E. H. (2002). Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen reading and prostate cancer. JASA 97(457), 53-65.

20. van der Laan M. J., Hubbard A. E., and Robin J.M. (2001). Locally efficient estimation of a multivariate survival function in longitudinal studies.

21. Dobson A., and Henderson R. (2003). Diagnostic for joint longitudinal and dropout time modeling. Biometrics 59, 741-751.

22. Henderson R., Diggle P.J., and Dobson A. (2002). Identification and efficacy of longitudinal markers for survival. Biostatistics 3(1), 33-50.

23. Klein, J.P. and Moeschberger M. L. (1997). Survival Analysis. New York: Springer-Verlag.

24. Klein, J.P. and Moeschberger M. L. (2003). Survival Analysis, Second Edition. New York: Springer-Verlag.

25. Fleming T. R., and Harrington D.P. (1991). Counting process and survival analysis. New York: Wiley.

26. Hougaard P. (2000). Analysis of multivariate survival data. New York: Springer-Verlag.

27. Cox D.R. and Oakes D. (1984). Analysis of Survival Data. Landon: Chapman & Hall.

28. Le C.T. (1997). Applied Survival Analysis. New York: Wiley.

29. Hosmer D.W., Jr and Lemeshow S. (1999). Applied Survival Analysis: Regression modeling of Time to Event Data. New York: Wiley.

30. McCullagh P. and Nelder J.A. (1989). Generalized Linear Models, Second edition. Landon: Chapman & Hall.

31. Hardin J. and Hilbe J (2001). Generalized Linear Models and Extensions. College Station: STATA Press.

32. Jones R.H. (1993). Longitudinal Data with Serial Correlation: A State-space Approach. Landon: Chapman & Hall.

33. Diggle P.J., Heagerty P., Liang K.Y. and Zeger S.L. (2002). Analysis of Longitudinal Data. New York: OXFORD University Press.

34. Hatteville L., Mahe C., and Hill C. (2002). Prediction of the long-term survival in breast cancer patients according to the present oncological status. Statistics in Medicine 21, 2345-2354.