**Time-dependent efficacy of longitudinal biomarker for clinical endpoint**

**Abstract:** Joint modelling of longitudinal biomarker and event-time processes has gained its popularity in recent years as they yield more accurate and precise estimates. Considering this modelling framework, a new methodology for evaluating the time-dependent efficacy of a longitudinal biomarker for clinical endpoint is proposed in this article. In particular, the proposed model assesses how well longitudinally repeated measurements of a biomarker over various time periods (0,t) distinguishes between individuals who developed the disease by time t and individuals who remain disease-free beyond time t. The receiver operating characteristic (ROC) curve is used to provide the corresponding efficacy summaries at various t based on the association between longitudinal biomarker trajectory and risk of clinical endpoint prior to each time point. The model also allows detecting the time period over which a biomarker should be monitored for its best discriminatory value. The proposed approach is evaluated through simulation and illustrated on the motivating dataset from a prospective observational study of biomarkers to diagnose the onset of sepsis.

**Keywords**: Joint modelling, Longitudinal biomarker, Time to event, ROC methodology, Sepsis

**Time-dependent efficacy of longitudinal biomarker for clinical endpoint**

**1. Introduction**

Many clinical and biomedical studies are aimed at discovering biomarkers that can accurately signal a clinical endpoint, e.g. measures of liver function such as pro-thrombin index as indicators of liver fibrosis [1], and in clinical practice, e.g. in critical care, rapid tests of biomarkers hold the promise of prompt diagnosis of diseases such as sepsis for an improved outcome [2]. In medical diagnostic research, receiver operating characteristics (ROC) curves are well established for assessing how well a biomarker is capable of discriminating between individuals who experience the disease onset (*cases*) from individuals who do not (*controls*) [3]. Often diagnostic accuracy study protocols are based on binary (*case*/*control*) disease outcome with a single biomarker measurement at baseline; however many disease outcomes and biomarkers are time dependent. Therefore, incorporating the time dimension in ROC analysis has recently been actively researched focusing on two factors: disease status is changing over time, and biomarker is measured longitudinally and varies with time. However, complications arise when event-time (e.g. time to disease onset) is censored and biomarker is measured intermittently. The current time-dependent ROC approaches deal with either a fixed (usually baseline) biomarker with censored event-time or address longitudinal nature of biomarkers with uncensored clinical endpoints [4, 5, 6-8]. However, unless information on longitudinal biomarker and censored event-time processes are combined correctly, the intermittently measured biomarker and missingness in measurement schedule could lead to misleading inference about the regression parameters that describe the true association between a prospective biomarker and subsequent risk of clinical endpoint. Further, the measurement error may hinder the explanatory power of the biomarkercausing the estimated parameters to be biased towards the null [9, 10]. In standard ROC methodology (where time is assumed stationary), Faraggi [11], Reiser [12] and others also concerned about ignoring the measurement error of biomarker measurements and showed that the effect can be substantial on the decision as to the diagnostic effectiveness of the biomarker.

The Cox proportional hazards (PH) regression model including the longitudinal biomarker as a time-varying covariate have achieved widespread use in the analysis of censored event-time outcomes in order to predict the risk of a clinical endpoint on a longitudinal biomarker (for example, Fisher and Lin [13]). However, Cox PH implementation requires that the longitudinal measurement is available at each event time while the measurement is collected on each individual intermittently at a discrete set of time points, and further, the observed biomarker measurement may not be the “true” values due to some measurement error. Therefore, this model must be used with caution [13-15]. A further difficulty for making inference on the longitudinal process is that occurrence of the event-time outcome may induce an informative censoring. For example, individuals in more serious condition may be more likely to experience the clinical endpoint than healthier individuals leading to fewer biomarker measurements, and sharper rates of decline in the biomarker over time [14].

Methodologies that jointly model both longitudinally repeated biomarker and censored event-time outcome processes have rapidly been developed in the past decade and as demonstrated in many studies these models combine the information from the two processes more efficiently (see Gould et al [16] and Tsiatis and Davidian [14] for comprehensive reviews of such modelling). Henderson et al. [17] and many others emphasised the importance of incorporating the complete biomarker information through a framework such as joint modelling in order to avoid biases due to intermittently measured biomarker, measurement error, and informative missingness in biomarker measurement schedule. However, there has been very little research on adopting this framework within the ROC curve analysis in estimating the discriminatory value of a longitudinal biomarker for clinical endpoint. Proust-Lima and Taylor [18] and Garre et al. [19] assumed a joint model with latent classes to explain the association between the longitudinal and event-time outcomes and developed a dynamic prognostic tool for early detection of the disease. Henderson et al. [20] has parameterised the underling association between longitudinal biomarker and event-time processes by individual-level deviation of the longitudinal profile from the population mean and derived a like statistic to quantify the predictive accuracy of a biomarker for a clinical endpoint. Rizopoulos [21] has assumed Gaussian latent effects with current value parameterisation and drawn accuracy summaries on predicted survival probabilities via the posterior expectation (an asymptotic Bayesian formulation) based on ROC methodology. In these two-stage approaches, the risk of clinical endpoint is predicted at certain time points by integrating the event-time distribution from the fitted joint model over the entire follow-up. These prediction approaches would be evidently valuable in everyday medical practice [21].

As its main contribution, this article provides a new development of evaluating the time-dependent effectiveness of a longitudinal biomarker for clinical endpoint by utilising a joint modelling framework. The proposed approach models the ROC curves associated with a longitudinally repeated biomarker at various time points given the risk of clinical endpoint prior to each time point. The model also allows detecting the time period over which a biomarker should be monitored for its best discriminatory value. Our proposal is motivated by a prospective observational study of sepsis biomarkers. Sepsis is a leading cause of death in critically ill patients and the diagnosis of sepsis is complicated by the highly variable symptoms of the disease [2]. It has been observed that the biomarker *activated partial thromboplastin time* (aPTT) was significantly more abnormal over early time periods among septic non-survivors than survivors [22]. This necessitates a rather different approach to assess the efficacy of aPTT over time for sepsis. We propose a novel approach to capture the variable impairment caused by the disease on a biomarker profile; we primarily focus on the change in underling association between the longitudinal biomarker and risk of disease over certain time periods within study follow-up since admission (at time ). In the proposed approach, firstly, a risk score for each individual is estimates at their event-time based on the joint (biomarker and event-time) information over , and secondly, the discriminatory potential of the estimated risk score for distinguishing between individuals who experienced the disease over from those who survived disease beyond time is computed within a ROC curve analysis. In particular, we determine the efficacy of the biomarker at each in terms of area under the ROC curve (AUC), sensitivity and specificity. We follow similar arguments (cumulative/dynamic) as Heagerty and colleagues [4, 5] in time-dependent ROC methodology to define cases and controls over each . The risk score over each time period is estimated within a joint modelling framework to allow for intermittently measured biomarker, measurement error, and informative missingness in biomarker measurement schedule.

The proposed approach is also compelling in many other studies in which diseases show highly variable symptoms over certain time frames and the interest resides in changes in longitudinal biomarker profiles for early diagnosis or disease progression to facilitate the immediate clinical management. We also use several simulation settings to validate the proposed joint modelling approach, and illustrate the methodology using the sepsis dataset from the study mentioned above.

**2. ROC curve analysis and integration of the time dimension**

**Traditionally in medical diagnostic research,** receiver operating characteristics (ROC) curve is used to assess the discrimination capability of a continuous measurement [3]. For a single biomarker measurement (usually the value at baseline) with a binary disease outcome indicator , an ROC curve analysis summarises

**for all possible values of , where indicates a *case* (disease present),**  **indicates a *control* (disease absent),** s estimates the expected fraction of individuals with (test positive)among *cases* (i.e. true positive fraction TPF), while estimates the expected fraction of individuals with (test negative)among controls (i.e. true negative fraction = false positive fraction FPF). When no a priori value of is indicated, the full spectrum of sensitivities and specificities can be characterised by ROC curve that plots the TPF versus the FPF for all . In the standard ROC curve analysis, the accuracy of is typically quantified with above two correct classification rates [3]. A perfect completely separates cases and controls; in that case, we have and , and then for any for a given threshold . If distributions for cases and controls are identical, hence is completely inefficient; then for any , we have , and therefore . The most familiar summary measure for the ROC function is the area under the ROC curve (AUC) which gives a single scalar value representing the expected performance of . AUC is the most often used ROC statistic for model comparison. AUC ranges from 0.5 to 1 for completely inefficient to perfect accuracy; See Pepe [23] for an excellent introduction to the standard ROC methodology.

**The time dimension can be included in the** standard ROC methodology by considering the time dependency in the biomarker measurement or binary disease outcome or both which **has recently been actively researched [4, 5, 6-8], and these approaches deal with either a fixed (baseline) biomarker with censored event-time outcome or address longitudinal nature of biomarkers with uncensored clinical endpoints.** Dealing with censored event-times in ROC curve analysis was discussed by Heagerty, Lumley and Pepe [4] based on the Kaplan-Meier survivor function and later by Heagerty and Zheng [5] and Zheng and Heagerty [8] based on the Cox regression model. The Kaplan-Meier and Cox regression are well established for modelling censored event-times, and hence the above approaches take proper account of censored event-times; however, a single biomarker value is considered usually at the baseline or at a certain time point (e.g. the final biomarker value measured prior to clinical endpoint); hence the longitudinal nature of biomarker measurements over has not been taken into account. A review of some methods based on Kaplan-Meier and Cox regression is given by Blanche *et al.* [24].

Etzioni et al. [6], and Zheng and Heagerty [7] discussed parametric and semi-parametric regression approaches characterising the longitudinal nature of a biomarker through linear mixed effect models. In their approaches, “time before event” is calculated at each biomarker measurement time and included it as a covariate in the model for *cases* (those who had failure) and it is considered irrelevant for *controls*, hence the censoring of the event-time process is not appropriately taken into account. Later, Zheng and Heagerty [8] discussed how well available measurement of the biomarker at time predicts an individual who becomes diseased in a subsequent time interval where , by extending Heagerty and Zheng [5] approach, but assume that any missingness in biomarker measurement schedule in this time period is noninformative (i.e. *missing completely at random* or MCAR, Rubin [25]). Although not implemented via an ROC methodology as above, de Bruijne et al [26] suggested using the biomarker value at event-time and “time elapsed since last biomarker measurement” as two time-varying covariates in the Cox PH model to predict the hazard on biomarker. The biomarker value at each event-time was estimated by the last observed value prior to event or using a function smoothing the observed biomarker history. This type of model is simpler to implement and fitting longitudinal outcomes as covariates in the Cox PH model has achieved widespread use in risk prediction, however, as previously mentioned, modelling biomarker as a time-varying covariate in the Cox PH model fails to account for the complete biomarker information on the subsequent event-time process [13-15]. The above model fails when the censoring process is informative on the longitudinal biomarker as identified by de Bruijne et al [26] themselves. The joint modelling of the longitudinal biomarker and event-time outcomes is a framework in which potential underlying relationship between the event-time and longitudinal process are explicitly acknowledged [e.g.9,10,15-17], and is implemented by updating the biomarker continuously throughout time as new biomarker measurements become available over time. Although, the joint modelling requires computationally intensive numerical integration, due to recent advances in computing and user-friendly software, such models can be implemented in real time.

**3. Joint modelling of longitudinal biomarker and clinical endpoint**

The joint modelling methodologies can be used to explicitly combine the information from **longitudinal** biomarker and censored event-time processes by modelling the hazard of the clinical endpoint at time conditional on biomarker history up to time .

**3.1 General notation**

Let be the failure time (e.g. disease onset) and denote the censoring time for the th individual. Let be the indicator of the event, therefore taking values 1 if the clinical endpoint is occurred at time , and 0 if it is not occurred (censored at ). We observe the event-time process where ) defines the time to event, and individuals in the study dataset. Denote to be a set of longitudinal response measurements for the th individual collected at times . We allow the possibility of different numbers and timing of longitudinal measurements for different individuals.

**3.2 Modelling formulation**

Usually, a joint model is formulated by two submodels; a longitudinal submodel for and an event-time submodel for ,, and the two components are linked together through some shared parameters. Longitudinal trajectory is typically modelled by linear mixed effect models, while the event-time assumes various choice of modelling approaches through shared latent effects [16]. The joint modelling formulation proposed by Henderson et al. [17] provides the appropriate specification for our purpose where we can model the underling association between longitudinal biomarker and event-time processes over by individual-level deviation of the longitudinal profile from the population mean Following Henderson et al [17], we assume a Gaussian linear model for biomarker measurement;

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

where and are regression coefficients related to intercept and slope. Individual-specific random effects are incorporated through where is an unobserved zero-mean Gaussian random processFinally, denotes the measurement error process and assumes Gaussian distribution with mean zero and variance . Further, and the measurement error process are assumed to be mutually independent [17]. We assume that the hazard for clinical endpoint at time is modelled by

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

where is the true unknown biomarker value at time , is an unspecified baseline hazard and is a second unobserved zero-mean Gaussian random processThe event-time process is associated with the longitudinal response through the shared random effect of and This model links the true biomarker value at time through the hazard of clinical endpoint at time for the th individual.

**4. Estimating the efficacy of a longitudinal marker**

Assume a collection of time points }since admission (at time 0) within the study follow-up. We focus on the change in underling association between the longitudinal biomarker and event-time process over , assuming that longitudinally recorded biomarker profiles over different time periods capture the variable impairment caused by disease over time. Our aim is to evaluate the time-dependency of the longitudinal biomarker on clinical endpoint by estimating how well the biomarker values measured over can distinguish between individuals who experienced the clinical endpoint by time and individuals who survived beyond time . The overlapping time periods since admission consider the biomarker over a longer follow-up at each time point.

We adopt the work by Heagerty and colleagues [4,5] on time-dependent ROC curves to define the *case/control* set at each . In the first stage of the proposed approach, a joint model of longitudinal biomarker process up to time and the event-time process related to the above *case/control* set is fitted and a risk score is estimated for each individual at their event-time over . In the second stage, we use the above estimated risk score to assess the discriminatory potential of the biomarker at empirically with respect to the observed proportions of *true* cases (experienced the clinical endpoint over by) and *true* controls (survived the clinical endpoint beyond ) conditional on a threshold value which determines the *test* positive and *test* negative (see section 2). The proposed approach also excludes the need for integrating an event-time distribution from the fitted joint model in order to predict the disease status at certain time points as in previously proposed studies [20,21]. This is an added advantage given already computationally intensive nature of the likelihood based EM (estimation-maximisation) estimation of joint modelling parameters [16]. We discuss the model and estimation process in detail below.

For simplicity in the notation, we use instead of in the rest of this section.

**4.1 Definitions**

At time , classify the individual in the entire study sample as either a *case* (experienced clinical endpoint by time) or a *control* (survived clinical endpoint beyond time ) on the basis of their event status at . Define an indicator such that each diseased individual () in the study sample plays a role as *control* for an early time but then play the role of *case* when . And each censored individual in the study sample () remains as *control*, hence if ; therefore

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

with indicates that th individual has experienced the clinical endpoint at time or prior to time and indicates that th individual has not experienced the clinical endpoint at time. Following Heagerty and colleagues [4,5], who has discussed a time-dependent ROC methodology for a single biomarker value with censored event-times, defines a cumulative/dynamic ROC terminology. It is the most appropriate indicator for our approach as for many other clinical studies as the scientific interest lies in discriminating between individuals who experience clinical endpoint prior to a given time and those that survive beyond [5].

Given at time and in study sample, we define an event-time for individual such that

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

That is, for each diseased individual in the study sample (), the event-time is kept at if and it is censored at time if and for each censored individual in the study sample (), the event time is set to , so the censoring time is brought forward at time if , and it is kept at if (censoring had occurred prior to time ). In other words, is an endpoint or horizon in its own time period . We then define the corresponding event-time process by . This is comparable to the corresponding counting process formulation at time [27].

Let be the observed longitudinal biomarker measurements for the th individual recorded over . This defines the observed history of biomarker measured over for each individual. We assume biomarker measurements are obtained intermittently at several time points over and are also measured with error. We specifically assume that occurrence of the time-to-event process induces an informative censoring; that is, missingness in biomarker measurements is *Missing Not at Random* (MNAR) [25], a more general form of missing data in many clinical studies.

Let define the observed joint outcome process for the th individual over , *.*

**4.2 Stage 1: Fitting the joint model**

We fit joint model by combining the information in the observed joint outcome process through longitudinal (1) and event-time (2) submodels as discussed in section 3.2. Let and be random intercept and random slope of the biomarker trajectory over () of the th individual and let assume a multivariate normal with mean 0 and variance [10]. and reflect individual-level deviations of the longitudinal profile from the population mean at baseline and from the population mean slope at event time respectively.

We assume a widely used association structure between the biomarker and clinical endpoint which is based on the individual-specific deviations from both the population mean intercept and population mean slope at event-time . This association structure is defined by where . We also explore the suitability of two further association structures in the supplementary file for implementing the proposed methodology.

In the above setting, submodel for the longitudinal biomarker over is determined by

and the hazard of clinical endpoint at time for individual is determined by event-time submodel

where is the true unknown longitudinal biomarker trajectory at the event-time . The parameter estimates the strength of the association between the current *true* biomarker measurement (at ) and occurrence of the clinical endpoint over . is also the log hazard ratio corresponding to the current biomarker measurement at time.

We estimate the association parameter by maximising the joint likelihood of the observed data via the EM algorithm and any integration is performed using the Gauss-Hermite quadrature. It involves taking expectations with respect to the unobserved latent process . The EM algorithm iterates between two steps until convergence is achieved. E-step determines expected values conditional on observed joint outcome . M-step maximises the complete data log-likelihood by replaced by corresponding expectation, and the association parameters are drawn. We also obtain the best linear unbiased estimates of the random effects [17, 20]. Once the joint model is implemented, we can calculate the corresponding model (or risk) score for the th individual at their event-time by

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

where , and is defined by (4) for a time period . Note that is estimated from the joint model at event-times for each individual, and it includes the weighted individual-specific deviation from true biomarker value at where weight is given by the estimated coefficient of the association parameter over .

**4.3 Stage 2: Cumulative sensitivity and dynamicspecificity over**

Let as the data over the time period is included in its estimation. Adopting the cumulative/dynamic definition by Heagerty and Zheng [5] for , we can classify for each individual as positive or negative for a given thresholding criterion ; defines a test positive and defines a test negative. Therefore, the sensitivity and specificity of the biomarker over can be defined by

where . estimates the fraction of individuals with among those who experience the clinical endpoint (disease onset) over , while estimates the fraction of individuals with among those who survived disease-free beyond time .

Provided that is estimated from the joint model for all individuals , the sensitivity and specificity can be empirically estimated using the distribution of separately for cases and controls. For a given thresholding criteria , we can calculate observed true positive fraction () by

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

and true negative fraction (1-) by

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

where is an indicator, and and respectively denote the “true” *case* over and the “true” *control* at and . A true *case* is an individual who experienced the clinical event by and a true *control* is an individual who survived event-free beyond . The denominators and respectively denote the sizes of the true *case* and *control* sets. The true *case/control* set is defined by removing all individuals censored before time from the entire study sample which may result in a loss of information in the estimation. However, as argued by Blanche *et al.* [24] through a comprehensive simulation study for a single biomarker value with event-time outcome, such empirical calculation gives reasonably accurate estimates with small biases in practice due to conditioning on actual event times. Further, in our approach, is an estimated model score from a joint modelling formulation accounting for an explicit dependence structure between the risk of clinical endpoint and biomarker profile over ; hence any bias induced by above empirical estimation is expected to be even smaller.

The ROC curve for all can be obtained by kernel (density) smoothing which follows closely the details of the original data [28]. The kernel estimate is obtained by smoothing the corresponding histograms of for the true cases and controls. Let the smoothed version of and be and respectively. The smoothed ROC function is given by for or equivalently

where is the abscissa axis () and the ordinate axis () of the ROC plot [28].

We can use similar calculations to predict the discriminatory potential of the biomarker at a future time based on over where . In this case, we consider the subsample who survived clinical endpoint at . By define the true *case/control* set similarly as above at using the current sample, ROC summaries can be similarly computed.

**4.4 Calculation of the 95% confidence intervals for sensitivity and specificity**

In this proposal, the estimated risk score is computed from joint modelling parameter estimates, which is then used as the input to ROC analysis. Hence the 95% confidence intervals (CIs) of sensitivity and specificity must account for uncertainty due to both estimation processes. Therefore, the 95% CIs for accuracy summaries are estimated by the bootstrap sampling with replacement [29]. The estimates are given by the corresponding bootstrap percentiles since the underlying distribution of the accuracies are unknown. The previously suggested time-dependent ROC models for censored clinical endpoint based on a single biomarker value were also suggested bootstrap approaches to estimate the corresponding CIs [4,5].

**5. Simulation study**

We have conducted several simulation studies to demonstrate whether the proposed approach is an appropriate framework for estimating the time-dependent accuracies of a longitudinal biomarker within ROC curve analysis. First, we explored the accuracy of estimation of association parameter from the joint model which is crucial for estimating the correct ROC summaries from the proposed approach. In second simulation study, we demonstrated how the strength of association between the biomarker and event-time process modifies the discriminatory potential of a biomarker (as estimated by AUC) as estimated by the proposed approach. Finally, in third simulation study, we evaluated whether the proposed estimation procedure is capable of identifying the time point at which the best discriminatory potential of the biomarker is reached.

We examined only negative associations ( to induce the direction of association in our application (it is anticipated that low aPTT values are associated with high risk of sepsis onset). However, the behaviour for positive associations with the same strength would be the same, but in opposite direction, and will not impact on biases and other characteristics.

**Simulation study 1:** The longitudinal values () were simulated for 500 individuals under a linear mixed model with fixed (population-level) intercept and slope with coefficients 1 and -1 respectively, and random intercept and random slope terms, and measurement error using The random intercept and random slope terms were generated from the bivariate normal distribution with variances 1 and covariance 0.5, and measurement error was generated from with variance . The positive covariance 0.5 and negative slope -1 induce a negative correlation between random effects and . This scenario stimulates an individual with a high biomarker value at the baseline to have a smaller drop in their biomarker values over time. Conversely, a lower marker value at baseline would lead to a more rapid decrease over time. Longitudinal times were set at 0, 1, 2, 3, 4, so a maximum of 5 longitudinal observations recorded at these time points up to individual’s event time in the final dataset.

Based on the association structure , event times were generated under Gompertz distribution with scale parameter and shape parameter assuming Cox proportional hazards model (see Bender et al. [30] for more details). The event times were simulated by

where is derived from the uniform[0,1] distribution, and . We set -3 and 1. Exponential distribution is used to control the censoring rate in the simulated data.

We varied {-0.25,-0.50,-0.75,-1} to allow weak (-0.25) to strong (-1) association between the biomarker and event-time outcomes. Exponential distribution parameter was set at to get the censoring rate approximately at 30%. We used 0.5, but set 1.5 and 2.5 to allow higher measurement error in biomarker values when is at the highest level to assess the impact of measurement error. We compared the estimated association parameter from the joint model against where is the most recent biomarker observation following Henderson *et al.* [20]. The regression parameter of the covariate indicates the association between risk of the event and biomarker; hence is comparable to in the joint modelling formulation with .

**Simulation study 2:** Longitudinal values () and event times () were generated as in simulation study 1 with 0.5, and a maximum of 5 longitudinal observations recorded at times 0, 1, 2, 3, 4 up to individual’s event time. We varied association parameter at {-0.25,-0.50,-0.75,-1} to explore how the strength of association modifies the discriminatory potential of the biomarker (as estimated by AUC). The % event was varied at 80, 50 and 30 by controlling the exponential distribution parameter for censoring. We estimated the association parameter and AUC from the proposed model for the entire sample. We have conducted similar simulation studies for two additional association structures as detailed in the supplementary file.

**Simulation study 3:** Assuming that the best discriminatory potential of a biomarker (as estimated by the highest AUC) should be estimated at , we placed a change point over the longitudinal profile at and set an association for the time frame pre- and for the time frame post- with , so a higher association between biomarker and event-time was set for the early time frame. Given the change point at , let and define an indicator 1 if and 0 otherwise. Longitudinal values were generated from with (population-level) slopes over both pre- and post- kept at -1 to allow the same scenario as in simulation study 1 and where is the random intercept term, and are the two random slope terms corresponding to pre- and post- time frames. The random effects were generated from the multivariate normal distribution with variances 1 and positive covariances 0.5. Based on the same association structure (split by ) , event times were generated by the same formulation with and . In this simulation, we generated longitudinal data over a longer time period at times = 0, 1, 2, 3, 4, 5, 6 to get enough follow-up values for later . We considered a setting with an approximately 30% events in the entire sample with overall association -0.5 to agree with the situation in application dataset. 0.5. The time point was varied at 2, 3 and 4. We followed the proposed estimation procedure to obtain the ROC summaries for time periods 1, 2, 3, 4, 5.

**5.1 Results from the simulation study 1**

Table 1 presents bias and coverage estimates for association parameter based on 10000 simulations. We observe that the joint model provides the most accurate estimation of the association with lower standard errors together with smaller biases and coverage probabilities closer to 95% across all settings, and this observation is consistent with the previous simulation study results [31]. Further, joint model estimates fairly close to the true value even when the measurement error is high; indicating that the joint model makes the proper adjustment of measurement error when estimating the underlying association. The Cox model with most recent biomarker measurement underestimates the association when the true dependence between the two outcomes is fairly strong, and high measurement error in longitudinal outcome substantially affects the estimation of association. The issues associated with the Cox model when estimating the association between a biomarker and event time are discussed previously in sections 1 and 2. These simulation results strengthened the case for using the joint model for estimating the efficacy of a longitudinal biomarker in practice.

**5.2 Results from the simulation study 2**

Table 2 tabulates the estimated and AUC from the proposed approach based on 500 simulations. When = 0, that is when there is no association between the biomarker and event-time process, the AUC is estimated fairly close to the null value of 0.5 which indicates biomarker shows no discriminatory potential. As strength of the association becomes stronger ( moves towards), the estimated AUC is increased by acceptable margins. Similar patterns can be found across different event rates although a slight drop in the estimated AUCs is observed as the event rate is decreased. We observe that even with a low event rate, the proposed methodology can comprehend the discriminatory potential of a biomarker (as estimated by AUC) when the corresponding association parameter is estimated correctly. The two additional association structures require a fairly high level of association between the biomarker and risk of clinical endpoint to be a useful form in the proposed methodology (see supplementary file for details).

**5.3 Results from the simulation study 3**

Table 3 tabulates the estimated ROC summaries based on 500 simulations for each . The average event rate over each time period is also shown. When 4, the ROC estimates for the time period could not be obtained due to considerably low event rate over that time period. The result from Table 3 indicates that the time at which the highest AUC was expected is correctly identified by the proposed estimation approach. For example, when was set at time 3 by simulation, the proposed approach estimated the best discriminatory potential of the biomarker over time period (0, 3). In other words, the proposed approach estimated the highest AUC at the correct time point in each setting. This observation confirms the validity of the proposed estimation process to identify the most appropriate time frame over which the biomarker should be monitored for the best discriminatory potential.

**6. Application: aPTT(TL18) as an early and time-dependent biomarker of sepsis**

We return to the motivating study introduced in Section 1. This is a single centre, prospective, observational study which includes patients admitted consecutively to the intensive treatment unit (ITU) of the Royal Liverpool University Hospital during January 2008 - June 2009. Sepsis is a major public health concern, and with a mortality rate of 30%, it emerges as a leading cause of death in critically ill patients. Early detection and timely therapeutic intervention is crucial for improved outcome of patients with sepsis [2]. One of the aims of this study is to explore whether aPTT can be an early and time-dependent biomarker of sepsis in critical care. aPTT has emerged as a promising biomarker for sepsis with a notable clinical potential over the traditional biomarkers [2,22]. aPTT is a standard test of coagulation, and its abnormality is quantified through the light transmittance level at 18 seconds (TL18). It is also identified as a simple, rapid and reproducible test, which fulfils practical requirements within the critical care setting [2]. aPTT is defined as abnormal if the TL18 was less than 99%. In this study, aPTT(TL18) measurements was recorded for each patient at least once daily since admission to the ITU. aPTT(TL18) values on different patients were not measured at a common set of time point, and not all patients had the same number of measurements over each time period; a typical unbalanced design observed in many longitudinal studies. All aPTT(TL18) measurements at the time they were measured since admission are allowed in the proposed model. We explore how the association between aPTT(TL18) and risk of sepsis onset varies over the critical care follow-up up to 7 days, aiming to expose whether longitudinal measurements of aPTT(TL18) facilitates early detection of sepsis.

Of the cohort, 129 patients were admitted to the ITU free of sepsis, and 39 (30%) developed sepsis during the first 7 days. Table 4 shows the number of *cases* and *controls* defined for where 24, 48, 72, 96, 120, 144, 168 hours. The dataset includes a total of 598 aPTT(TL18) records; of which 186 belonged to the 30% who had developed sepsis over follow-up. Figure 1(a) and (b) shows corresponding smooth mean profiles of aPTT(TL18) relative to event-time across all patients for those who developed sepsis during (cases, ) and for those who had been defined as free of sepsis at time (controls, ) respectively. The mean aPTT(TL18) profile drop significantly below 99% towards the time of sepsis diagnosis for cases, while they remain fairly stable closer to 99% for controls. Further, the drop in mean aPTT(TL18) profiles noticeably varies between earlier and later time periods since ITU admission among cases. This observation sketches the development of the proposed methodology; as such the time-dependent effectiveness of the biomarker is evaluated to account for the varying association between the longitudinal biomarker profile and *case/control* set over each time period .We analyse aPTT(TL18) in loge scale as longitudinal outcome required to be normally distributed for the proposed model. Since relatively fewer numbers of patients developed sepsis at later follow-up times, we consider 24, 48, 72, 96, 120 in this analysis. To describe the sharp drop in mean aPTT(TL18) profiles between time 0 and 24 hours (see Figure 1(a) and (b)), we include an indicator variable for measurements at time zero in the model.

Parameter estimates and standard errors (SE) from the fitted joint model are given in Table 5. We have also fitted two further joint models assuming the additional association structures as discussed in the supplementary file (see supplementary file for the model estimates). The maximised joint log-likelihood values and standard errors of the estimated parameters strongly favour latent association structure considered in the main article over the two additional structures. As given in Table 5, the estimated association parameters over different time periods indicate that there is significantly strong evidence of association between the aPTT(TL18) and risk of sepsis, and that association is particularly strong over earlier time periods compared to later time periods. Negative estimates of association parameters suggest that lower aPTT(TL18) are associated with higher risk of sepsis onset. The estimated measurement error for the longest period, i.e. (0, 120), 0.315 (SE 0.039) indicates a significantly high level of error or noise in aPTT(TL18) measurements.

The estimated ROC summaries and SEs from the proposed approach are given in Table 6 together with estimates from three currently used time-dependent ROC methodologies for comparison. Some details of the three approaches are also discussed in section 2. The SEs for all four models were estimated from 500 bootstrap resamples. Heagerty et al (2000)[4] defined the event-time process over each time period similar to our approach, and estimated the time-dependent ROC summaries at from censored event-time using the Nearest Neighbor Estimation (NNE) method based on baseline biomarker measurement. This model is popular among clinical applications in medical diagnostic research (from an unpublished review), and can be fitted using the survivalROC library in R language. The estimated AUC() from this approach is quite close to that from the proposed model at = 24 hours due to similarities in event-time processes. The AUC()s from Heagerty et al (2000)[4] decrease as increases since in general, the predictive ability of the biomarker at baseline is dropped over time; however, in our approach, AUC() is estimated on the association between the biomarker and sepsis risk over each time period while the biomarker trajectory is updated continuously over time.

Heagerty and Zheng (2005)[5] also used the baseline biomarker measurement to estimate time-dependent ROC summaries at based on Cox regression and risksetROC library in R is available to fit this model. This approach considers the event-times that occurred exactly at only (ignored events occurred between 0 and ). Etzioni et al (1999)[6] defined the time-dependent ROC summaries at based on a liner mixed effect model for longitudinal biomarker, similar to the longitudinal sub-model in the proposed joint model, however, as discussed earlier in section 2, the model included *time before event* as a covariate to account for the event-time of those individuals who had sepsis (cases), but ignored the censored event-times. The estimated AUC() from this model is similar to Heagerty and Zheng (2005)[5] at early time points as the model included information for exact event times for cases; however, as additional (longitudinal) biomarker values were included, the estimated AUC()s became higher for later than those from Heagerty and Zheng (2005)[5]. Our approach jointly models the event-time process and the longitudinal biomarker measurements prior to each and therefore considered a more complete record of each patient when estimating the ROC summaries over time.

As shown in Table 6, the highest AUC from the proposed approach is estimated at 24 hours, which is 72.2% and implies that the probability of a patient who developed sepsis within 24 hours following ITU admissionhas a significantly lower aPTT(TL18) value than a patient who remained sepsis-free beyond 24 hoursis at least 72.2%. The corresponding estimate of sensitivity is 75% at the optimal threshold indicates that the measurements of aPTT(TL18) over the first 24 hours since ITU admission can be used to detect 75% of septic cases correctly. The estimated AUC, sensitivity and specificity at 48 hours are also remained close to that at 24 hours. Figure 2 shows the corresponding ROC curves for the occurrence of sepsis at times 24, 48, 72, 96, 120 hours from the proposed approach. The predictive discriminatory value of aPTT(TL18) at a future time point based on the (current) estimates of over is shown in Figure 3. The predictive value of the model is high at 48 hours, but dropped as is getting further away from = 24 as expected.

From this analysis, we observed that longitudinally measured aPTT(TL18) alone could be a useful biomarker for ruling-out sepsis within 48 hours following admission to the ITU. Further, estimates of association parameter over each time period provide indices for time-dependent nature of aPTT(TL18) for sepsis onset which can be used for improved patient management (e.g. to assess progression or treatment responses). Although the sample size in this application is relatively small, the jointly modelling of multiple dimensions of data yields a greater precision; greater precision implies higher power and smaller sample sizes. Incorporating the longitudinal data in a study has the potential of yielding lower sample sizes with higher power compared with that of conventional study based on time-to-event data alone [32]. However, we may need a larger sample of patients to make stronger recommendations for the potential use of aPTT(TL18) in sepsis management in the intensive care.

**7. Discussion**

In this work we proposed a two-stage approach within a joint modelling framework in an attempt to estimate the time-dependent efficacy of a longitudinal biomarker for clinical endpoint by focusing on the change in association between the longitudinal biomarker and the corresponding event-time processes over time periods () since admission. We have shown that joint modelling formulation accounts for the underlying association between longitudinal biomarker (measured with error) and the risk of subsequent clinical endpoint (censored) explicitly, and hence can yield the most precise inferences concerning biomarker effectiveness at clinically defined time points. The proposed formulation is built upon familiar, more traditional theory of ROC curve analysis, and is implemented in real time using the existing statistical software libraries in R language. We have used the joineR library [33] in R language to estimate the joint model parameters, however, a variety of software with alternative estimation procedures are available for this class of models including the Bayesian approaches (e.g. JM[34] and JMbayes [35] in R, and stjm in STATA [36]). The R code is available on request from the corresponding author.

The proposed approach can also readily be used to summarise whether the efficacy of the longitudinal biomarker varies across patient subgroups by including the relevant covariates, e.g. age group or treatment arm, when estimating the risk score at Stage 1. The suggested approach can also be extended to allow competing risks for disease onset [37] by altering the definition of . An added distinction of the proposed approach is that, sequentially estimated association parameters over meaningfully (clinically) spaced time periods since entry can be used to provide indices of progression or treatment responses and thereby to provide information that can direct choice of therapy.

Based on current evidence, most utilised biomarkers for sepsis lack the necessary accuracy to be used without clinical judgement [22]. We believe that the accuracy of the proposed aPTT(TL18) model can be improved further by identifying baseline demographic and clinical characteristics of patients that interact the response (time of sepsis onset as well as aPTT), e.g. age, and the choice of threshold. If such characteristics and threshold were identified on adequate sample size, then the final model would also need validation in an external cohort before being taken further into a randomised control trial to evaluate an early intervention strategy guided by aPTT(TL18) model over the first 48 hours following ITU admission. Although, such work is beyond the scope of this paper, we illustrated that the proposed methodology can be used to guide detection and management of sepsis more accurately.

**Competing interests**

The authors declare that they have no competing interests.

**Acknowledgements**

The authors thank Professors Cheng-Hock Toh and Ingeborg Welters, and Mr Colin Downey from the Roald Dahl Haemostasis and Thrombosis Centre, Royal Liverpool University Hospital, Liverpool, UK for providing the aPTT biomarker data for the application. We also thank reviewers for their constructive and helpful comments.

**References**

1. Adams LA. Biomarkers of liver fibrosis. J. Gastroen Hepatol. 2011; 26(5): 802-809.

2. Toh CH, Ticknor LO, Downey C, Giles AR, Paton R, Wenstone R. Early identification of sepsis and mortality risks through simple, rapid clot-waveform analysis. Intens. Care Med. 2003; 29: 55-61.

3. Zweig MH, Campbell G. Receiver-operator characteristic plots: A fundamental evaluation tool in clinical medicine. Clin. Chem. 1993; 39: 561–577.

4. Heagerty PJ, Lumley T, Pepe MS. Time dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000; 56:337-344.

5. Heagerty PJ, Zheng Y. Survival Model Predictive Accuracy and ROC curves. Biometrics. 2005; 61: 92 – 105.

6. Etzioni R, Pepe M, Longton G, Hu C, Goodman G. Incorporating the time dimension in receiver operating characteristic curves: a case study of prostate cancer. Med Decis Making. 1999; 19: 242–251.

7. Zheng Y, Heagerty PJ. Semiparametric estimation of time-dependent ROC curves for longitudinal marker data. Biostatistics. 2004; 5: 615–632.

8. Zheng Y, Heagerty P. Prospective accuracy for longitudinal markers. Biometrics. 2007; 63: 332–341.

9. Tsiatis AA, Degruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. J. Am. Stat. Assoc. 1995 ; 90: 27–37.

10. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. Biometrics. 1997; 53: 330–339.

11. Faraggi D. The effect of random measurement error on receiver operating characteristic (ROC) curves. Statist. Med. 2000; 19: 61–70.

12. Reiser B. Measuring the effectiveness of diagnostic markers in the presence of measurement error through the use of ROC curves. Statist. Med. 2000; 19: 2115–2129.

13. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. Annu Rev Public Health. 1999;20:145-57.

14. Tsiatis A, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. Statistica Sinica. 2004; 14: 809–34.

15. Sweeting MJ, Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. Biometrical Journal. 2011; 53:750–763.

16. Gould AL, Boye ME, Crowther MJ, Ibrahim JG, Quartey G, Micallef S, Bois FY. Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. Statistics in Medicine 2015;34(14):2181-2195.

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

18. Proust-Lima C, Taylor J. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of post treatment PSA: A joint modeling approach.

Biostatistics. 2009; 10, 535–549.

19. Garre F, Zwinderman A, Geskus R, Sijpkens Y. A joint latent class changepoint model to improve the prediction of time to graft failure. Journal of the Royal Statistical Society, Series A, 2008; 171, 299–308.

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

21. Rizopoulos, D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. Biometric. 2011; 67(3): 819-829.

22. Chopin N, Floccard B, Sobas F et al. Activated partial thromboplastin time waveform analysis: a new tool to detect infection? Crit Care Med 2006; 34: 1654–60.

23. Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. New York, Oxford University Press. 2003.

24. Blanche P, Dartigues JF, Jacqmin-Gadda H. Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring. Biom. J. 2013; 55(5): 687-704.

25. Rubin DB. Inference and missing data (with discussion). Biometrika. 1976; 63: 581–592.

26. de Bruijne MHJ, le Cessie S, Kluin-Nelemans HC, van Houwelingen HC. On the use of Cox regression in the presence of an irregularly observed time-dependent covariate. Statist. Med. 2001; 20: 3817–3829.

27. Hosmer DW, Lemeshow S. Applied Survival Analysis: Regression Modelling of Time to Event Data. New York: Wiley.1999.

28. Zou KH, Hall WJ, Shapiro DE. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. Stat Med. 1997;16:2143–2156.

29. Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman and Hall.1993.

30. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. Statist. Med. 2005; 24: 1713-1723.

31. Crowther MJ, Lambert PC, Abrams KR. Adjusting for measurement error in baseline prognostic biomarkers included in a time-to-event analysis: a joint modelling approach. BMC Medical Research Methodology. 2013; 13:146.

32. Ibrahim JG, Chu H, Chen LM. Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data. ‎J. Clin. Oncol. 2010; 28 (16): 2796-2801.

33. Philipson P, Diggle P, Sousa I, Kolamunnage-Dona R, Williamson P, Henderson R. [joineR: Joint modelling of repeated measurements and time-to-event data.](http://nrl.northumbria.ac.uk/6945/) 2012. *Comprehensive R Archive Network*, United Kingdom.

34. Rizopoulos D. JM: an R package for the joint modelling of longitudinal and time-to-event data. J Stat Soft 2010;35: 1–33.

35. Rizopoulos D. JMbayesL joint modeling of longitudinal and time-to-event data under a Bayesian approach, 2013. Available from: http://cran.r-project.org/web/packages/JMbayes/index.html.

36. Crowther MJ, Abrams KR, Lambert PC. Joint modeling of longitudinal and survival data. The STATA Journal 2013; 13:165–184.

37. Williamson PR, Kolamunnage-Dona R, Philipson P, Marson AG. Joint modelling of longitudinal and competing risks data. Statist. Med. 2008; 27: 6426-6438.

Table 1. Simulation results of bias and coverage of association parameters for varying measurement error.

|  |  |  |
| --- | --- | --- |
| **True association parameter** | **Joint model** | **Cox regression model with the most recent measurement** |
|  | **SE** | **Bias** | **Coverage** |  | **SE** | **Bias** | **Coverage** |
| Measurement error 0.5 |
| -0.25 | -0.248 | 0.024 | 0.002 | 95.1 | -0.297 | 0.023 | -0.047 | 46.8 |
| -0.5 | -0.495 | 0.035 | 0.005 | 94.7 | -0.498 | 0.030 | 0.002 | 94.8 |
| -0.75 | -0.734 | 0.050 | 0.016 | 93.9 | -0.647 | 0.036 | 0.103 | 18.1 |
| -1 | -0.961 | 0.067 | 0.039 | 91.4 | -0.757 | 0.041 | 0.243 | 0.0 |
| Measurement error 1.5 |
| -1 | -0.912 | 0.088 | 0.088 | 83.8 | -0.535 | 0.032 | 0.465 | 0.0 |
| Measurement error 2.5 |  |
| -1 | -0.873 | 0.098 | 0.127 | 73.4 | -0.428 | 0.027 | 0.572 | 0.0 |

Table 2. Simulated results of varying strength of association between the biomarker and event-time process and discriminatory potential of the biomarker (as estimated by area under the receiver operating characteristic curve AUC)

|  |  |  |
| --- | --- | --- |
| **True**  |  | **AUC** |
| **Est.** | **SE** | **Est.** | **SE** |
| **80% events** |
| 0 | -0.009 | 0.006 | 0.518 | 0.019 |
| -0.25 | -0.251 | 0.020 | 0.586 | 0.029 |
| -0.5 | -0.494 | 0.024 | 0.694 | 0.028 |
| -0.75 | -0.738 | 0.026 | 0.774 | 0.025 |
| -1 | -0.956 | 0.040 | 0.860 | 0.021 |
| **50% events** |
| 0 | -0.009 | 0.005 | 0.517 | 0.016 |
| -0.25 | -0.249 | 0.022 | 0.582 | 0.025 |
| -0.5 | -0.496 | 0.027 | 0.671 | 0.023 |
| -0.75 | -0.744 | 0.026 | 0.750 | 0.022 |
| -1 | -0.978 | 0.050 | 0.824 | 0.022 |
| **30% events** |
| 0 | -0.009 | 0.006 | 0.515 | 0.020 |
| -0.25 | -0.248 | 0.025 | 0.569 | 0.027 |
| -0.5 | -0.500 | 0.028 | 0.649 | 0.025 |
| -0.75 | -0.745 | 0.029 | 0.749 | 0.024 |
| -1 | -0.986 | 0.056 | 0.807 | 0.022 |
|  |

Table 3. Simulated results for estimated ROC summaries from the proposed approach for time points at which the best discriminatory potential of the biomarker is reached (as estimated by the highest AUC)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Time period** | **% events** |  | **AUC** | **Sensitivity** | **Specificity** |
| **Est.** | **SE** | **Est.** | **SE** | **Est.** | **SE** | **Est.** | **SE** |
|  **2** |
| (0, 1) | 10.8 | -0.391 | 0.112 | 0.660 | 0.045 | 0.689 | 0.137 | 0.597 | 0.135 |
| **(0, 2)** | **20.5** | **-0.626** | **0.083** | **0.735** | **0.030** | **0.735** | **0.085** | **0.646** | **0.086** |
| (0, 3) | 27.5 | -0.573 | 0.113 | 0.711 | 0.038 | 0.713 | 0.097 | 0.634 | 0.098 |
| (0, 4) | 31.2 | -0.568 | 0.057 | 0.703 | 0.026 | 0.682 | 0.084 | 0.653 | 0.089 |
| (0, 5) | 32.5 | -0.520 | 0.039 | 0.677 | 0.024 | 0.638 | 0.090 | 0.665 | 0.092 |
|  **3** |
| (0, 1) | 3.1 | -0.217 | 0.113 | 0.601 | 0.071 | 0.675 | 0.180 | 0.627 | 0.176 |
| (0, 2) | 12.3 | -0.323 | 0.073 | 0.603 | 0.033 | 0.729 | 0.124 | 0.529 | 0.126 |
| **(0, 3)** | **24.6** | **-0.493** | **0.041** | **0.637** | **0.022** | **0.816** | **0.071** | **0.464** | **0.076** |
| (0, 4) | 29.8 | -0.462 | 0.030 | 0.584 | 0.022 | 0.780 | 0.089 | 0.418 | 0.093 |
| (0, 5) | 30.4 | -0.460 | 0.029 | 0.570 | 0.022 | 0.755 | 0.091 | 0.423 | 0.095 |
|  **4** |
| (0, 1) | 0.2 | -0.021 | 0.128 | - | - | - | - | - | - |
| (0, 2) | 1.4 | -0.053 | 0.136 | 0.595 | 0.081 | 0.723 | 0.200 | 0.596 | 0.202 |
| (0, 3) | 5.9 | -0.172 | 0.109 | 0.587 | 0.047 | 0.759 | 0.150 | 0.473 | 0.155 |
| **(0, 4)** | **17.3** | **-0.502** | **0.066** | **0.626** | **0.026** | **0.819** | **0.078** | **0.465** | **0.086** |
| (0, 5) | 27.7 | -0.402 | 0.039 | 0.583 | 0.022 | 0.809 | 0.073 | 0.390 | 0.076 |

Table 4. Number of cases (developed sepsis over ) and controls (remained sepsis-free at ) over study follow-up

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  **(hours)** | **0** | **24** | **48** | **72** | **96** | **120** | **144** | **168** |
| N Cases | 0 | 17 | 22 | 25 | 29 | 35 | 37 | 39 |
| N Controls | 129 | 112 | 107 | 104 | 100 | 94 | 92 | 90 |

Table 5. Sepsis study results from the joint model over time periods where = 24, 48, 72, 96 hours

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  **= 24** |  **= 48** |  **= 72** |  **= 96** |  **= 120** |
| **Est** | **SE** | **Est** | **SE** | **Est** | **SE** | **Est** | **SE** | **Est** | **SE** |
| Constant | -1.656 | 0.140 | -1.713 | 0.128 | -1.949 | 0.133 | -1.999 | 0.124 | -1.981 | 0.120 |
| Time,  | -0.008 | 0.009 | -0.006 | 0.004 | 0.006 | 0.003 | 0.008 | 0.002 | 0.008 | 0.002 |
| Indicator, = 0 | 0.208 | 0.096 | 0.268 | 0.089 | 0.505 | 0.087 | 0.555 | 0.094 | 0.538 | 0.093 |
| Latent association,  | -0.556 | 0.246 | -0.554 | 0.262 | -0.493 | 0.210 | -0.440 | 0.206 | -0.466 | 0.182 |
|  | 1.027 | 0.099 | 0.933 | 0.110 | 0.956 | 0.119 | 0.940 | 0.111 | 0.915 | 0.104 |
|  | -0.004 | 0.000 | -0.003 | 0.000 | -0.005 | 0.000 | -0.004 | 0.000 | -0.003 | 0.000 |
|  | 0.002 | 0.005 | 0.001 | 0.003 | 0.000 | 0.003 | 0.000 | 0.002 | 0.000 | 0.001 |
| Measurement error | 0.052 | 0.018 | 0.203 | 0.043 | 0.253 | 0.032 | 0.306 | 0.038 | 0.315 | 0.039 |
| Joint log-likelihood | -372.504 | -524.440 | -621.848 | -715.838 | -791.582 |

Table 6. Estimated AUC(), sensitivity() and specificity() at = 24, 48, 72, 96, 120 hours from the proposed and alternative approaches

|  |  |  |  |
| --- | --- | --- | --- |
|  | **AUC (SE)** | **Sensitivity (SE)**‡ | **Specificity (SE)**‡ |
| **Proposed approach** |
| 24 | 0.722 (0.077) | 0.750 (0.110) | 0.584 (0.119) |
| 48 | 0.699 (0.069) | 0.753 (0.083) | 0.547 (0.103) |
| 72 | 0.683 (0.057) | 0.746 (0.067) | 0.531 (0.082) |
| 96 | 0.666 (0.061) | 0.771 (0.075) | 0.490 (0.089) |
| 120 | 0.674 (0.068) | 0.744 (0.084) | 0.520 (0.106) |
| **Heagerty et al (2000)[4]** |
| 24 | 0.728 (0.056) | 0.725 (0.085) | 0.630 (0.071) |
| 48 | 0.714 (0.057) | 0.687 (0.078) | 0.633 (0.073) |
| 72 | 0.670 (0.057) | 0.673 (0.083) | 0.574 (0.080) |
| 96 | 0.649 (0.069) | 0.681 (0.086) | 0.583 (0.061) |
| 120 | 0.644 (0.062) | 0.613 (0.095) | 0.593 (0.083) |
| **Heagerty and Zheng (2005)[5]** |
| 24 | 0.646 (0.045) | 0.608 (0.040) | 0.622 (0.037) |
| 48 | 0.618 (0.051) | 0.585 (0.041) | 0.597 (0.042) |
| 72 | 0.598 (0.056) | 0.576 (0.046) | 0.576 (0.046) |
| 96 | 0.562 (0.048) | 0.550 (0.037) | 0.545 (0.041) |
| 120 | 0.561 (0.057) | 0.545 (0.043) | 0.548 (0.048) |
| **Etzioni et al (1999)[6]**  |
| 24 | 0.629 (0.073) | 0.595 (0.063) | 0.589 (0.062) |
| 48 | 0.636 (0.071) | 0.590 (0.060) | 0.605 (0.064) |
| 72 | 0.643 (0.076) | 0.593 (0.061) | 0.612 (0.071) |
| 96 | 0.649 (0.086) | 0.594 (0.067) | 0.620 (0.080) |
| 120 | 0.653 (0.100) | 0.600 (0.077) | 0.620 (0.091) |

‡ *Sensitivities and specificities are derived at the optimum probability threshold.*



Figure 1. Observed mean profiles of aPTT LT18 measured over (0,t) for (a) cases (had sepsis onset prior to or at time t) and (b) controls (free of sepsis at time t) where t = 24, 48, 72, 96, 120, 144, 168 hours.



Figure 2. Estimated ROC(t) curves at times t = 24, 48, 72, 96 and 120 hours



Figure 3. Predicted ROC(t\*) curves at future times t\* = 48, 72, 96 and 120 hours from t = 24 hours

**Supplementary File: Additional association structures**

**(1) Model A: (Model A)**

This association structure assumes that the association between the biomarker and clinical event is based only on the individual-specific deviations from the population mean intercept, that is, baseline value. The joint model is defined by

where is the true unknown longitudinal biomarker trajectory at the event-time .The parameter estimates the strength of the association between the individual-specific baseline biomarker measurements and occurrence of the clinical endpoint over , and is the hazard ratio for a one-unit increase in above deviation at the baseline. Therefore, in this model, the baseline biomarker value alone is linked to hazard for occurrence of the clinical endpoint over . In this model, the latent association is time independent.

**(2) Model C: (Model C)**

This association structure assumes that the individual-specific deviations from the population mean intercept and slope of the corresponding biomarker profile to have separate effects on the clinical endpoint. The joint model is defined by

where and estimate the strength of the association between the biomarker at baseline and at time and occurrence of the clinical endpoint over respectively. In this model

Once the joint model is implemented, we calculate the corresponding model (or risk) score for the th individual at their event-time by

|  |  |  |
| --- | --- | --- |
|  |  |  |

where and

where is defined by (4) (see main article) for a time period , and estimates the corresponding association over .

**Simulation study**

We have conducted the simulation study 2 based on the above two association structures to explore how these association structures affect the discriminatory potential of a biomarker (as estimated by AUC). Event rates were varied at 80% and 30%.

**Table S1. Model A**

|  |  |  |
| --- | --- | --- |
| **True**  |  | **AUC** |
| **Est.** | **SE** | **Est.** | **SE** |
| **80% events** |
| 0 | -0.010 | 0.006 | 0.520 | 0.019 |
| -0.25 | -0.236 | 0.025 | 0.550 | 0.026 |
| -0.5 | -0.492 | 0.028 | 0.590 | 0.023 |
| -0.75 | -0.749 | 0.028 | 0.633 | 0.025 |
| -1 | -0.979 | 0.053 | 0.678 | 0.027 |
| **30% events** |
| 0 | -0.010 | 0.006 | 0.517 | 0.019 |
| -0.25 | -0.239 | 0.027 | 0.540 | 0.030 |
| -0.5 | -0.497 | 0.028 | 0.569 | 0.031 |
| -0.75 | -0.748 | 0.027 | 0.593 | 0.031 |
| -1 | -0.990 | 0.054 | 0.627 | 0.025 |
|  |

As given by Table S1, when there is no association between the biomarker and event-time process, the AUC is estimated fairly close to the null value of 0.5 from the Model A. As strength of the association becomes stronger, the estimated AUC is increased; however, the model only links the baseline value of the biomarker to hazard for occurrence of the clinical endpoint ignoring any effect from random slope over time. Therefore, unless the association is very high (possibly ), the model may not able to detect the discriminatory potential of the biomarker. This observation has not been changed on the event rate. This model may only be useful if there is fairly high association between the biomarker and risk of clinical endpoint.

For the association structure in model C, we have considered several combinations for event rates at 80% and 30%. The results from the simulation are tabulated in Table S2.

**Table S2. Model C**

|  |  |  |  |
| --- | --- | --- | --- |
| **True**  |  |  | **AUC** |
| **Est.** | **SE** | **Est.** | **SE** | **Est.** | **SE** |
| **80% events** |  |  |
| -0.25; -1 | -0.255 | 0.028 | -0.924 | 0.173 | 0.737 | 0.032 |
| -0.5; -1 | -0.497 | 0.027 | -0.939 | 0.178 | 0.703 | 0.032 |
| -0.5; -0.25 | -0.500 | 0.029 | -0.270 | 0.082 | 0.579 | 0.027 |
| **30% events** |  |
| -0.25; -1 | -0.251 | 0.029 | -0.897 | 0.103 | 0.694 | 0.026 |
| -0.5; -1 | -0.497 | 0.028 | -0.891 | 0.119 | 0.650 | 0.026 |
| -0.5; -0.25 | -0.499 | 0.027 | -0.253 | 0.050 | 0.556 | 0.029 |

Model C allows baseline and slope of the biomarker profile to have separate effects on the clinical endpoint. As shown in Table S1, when either (at baseline or at event-time) is fairly strong, the estimated AUC is appropriately high. However, this model may not be useful if both and are weaker, and requires longer computational time due to the increased complexity from separate association parameters. Further work is needed to validate this model via the proposed estimation procedure for identifying the time at which the best discriminatory potential of the biomarker is reached.

**Table S3.** Sepsis study results from the above latent association models over time periods where = 24, 48, 72, 96 hours

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  **= 24** |  **= 48** |  **= 72** |  **= 96** |  **= 120** |
| **Est** | **SE** | **Est** | **SE** | **Est** | **SE** | **Est** | **SE** | **Est** | **SE** |
| **Model A** |  |  |  |  |  |
| Constant | -1.620 | 0.181 | -1.707 | 0.128 | -1.925 | 0.126 | -1.984 | 0.128 | -1.972 | 0.119 |
| Time,  | -0.010 | 0.012 | -0.006 | 0.004 | 0.006 | 0.003 | 0.008 | 0.002 | 0.008 | 0.002 |
| Indicator, = 0 | 0.171 | 0.154 | 0.260 | 0.092 | 0.479 | 0.091 | 0.538 | 0.100 | 0.528 | 0.096 |
| Latent association,  | -0.693 | 0.291 | -0.632 | 0.253 | -0.540 | 0.235 | -0.490 | 0.219 | -0.518 | 0.186 |
|  | 0.898 | 0.112 | 0.889 | 0.097 | 0.823 | 0.095 | 0.780 | 0.093 | 0.777 | 0.093 |
| Measurement error | 0.287 | 0.052 | 0.351 | 0.045 | 0.419 | 0.051 | 0.427 | 0.047 | 0.414 | 0.044 |
| Joint log-likelihood | -393.622 | -536.886 | -644.654 | -734.611 | -809.629 |
| **Model C** |  |  |  |
| Constant | -1.660 | 0.143 | -1.746 | 0.124 | -1.956 | 0.125 | -2.004 | 0.128 | -1.986 | 0.119 |
| Time,  | -0.007 | 0.009 | -0.003 | 0.004 | 0.006 | 0.002 | 0.008 | 0.002 | 0.008 | 0.001 |
| Indicator, = 0 | 0.209 | 0.100 | 0.293 | 0.088 | 0.509 | 0.087 | 0.558 | 0.092 | 0.541 | 0.091 |
| Latent association,  | -0.586 | 0.250 | -0.565 | 0.276 | -0.500 | 0.227 | -0.449 | 0.197 | -0.488 | 0.183 |
| Latent association,  | 0.182 | 0.138 | 1.135 | 0.501 | -0.221 | 0.632 | -0.167 | 0.699 | -0.235 | 0.496 |
|  | 1.029 | 0.102 | 0.943 | 0.107 | 0.955 | 0.119 | 0.939 | 0.106 | 0.915 | 0.102 |
|  | -0.005 | 0.000 | -0.005 | 0.000 | -0.005 | 0.000 | -0.004 | 0.000 | -0.003 | 0.000 |
|  | 0.002 | 0.005 | 0.001 | 0.003 | 0.000 | 0.003 | 0.000 | 0.002 | 0.000 | 0.001 |
| Measurement error | 0.051 | 0.018 | 0.200 | 0.040 | 0.253 | 0.030 | 0.305 | 0.037 | 0.315 | 0.040 |
| Joint log-likelihood | -372.043 | -523.205 | -621.769 | -715.731 | -791.439 |