

A Bayesian model for misclassified binary outcomes and correlated survival data with applications to breast cancer^{‡,§}

Sheng Luo,^{a,*†} Min Yi,^b Xuelin Huang^c and Kelly K. Hunt^b

Breast cancer patients may experience ipsilateral breast tumor relapse (IBTR) after breast conservation therapy. IBTR is classified as either true local recurrence or new ipsilateral primary tumor. The correct classification of IBTR status has significant implications in therapeutic decision-making and patient management. However, the diagnostic tests to classify IBTR are imperfect and prone to misclassification. In addition, some observed survival data (e.g., time to relapse, time from relapse to death) are strongly correlated with IBTR status. We present a Bayesian approach to model the potentially misclassified IBTR status and the correlated survival information. We conduct the inference using a Bayesian framework via Markov chain Monte Carlo simulation implemented in WinBUGS. Extensive simulation shows that the proposed method corrects biases and provides more efficient estimates for the covariate effects on the probability of IBTR and the diagnostic test accuracy. Moreover, our method provides useful subject-specific patient prognostic information. Our method is motivated by, and applied to, a dataset of 397 breast cancer patients. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: binomial regression; Cox model; frailty model; latent class model; Markov chain Monte Carlo; tumor relapse

1. Introduction

Breast conservation therapy (BCT) offers similar overall survival and disease-free survival rates as does mastectomy in breast cancer patients [1]. Approximately 8–20% of patients experience ipsilateral breast tumor relapse (IBTR), defined as the recurrence of tumor in the previously treated breast, after undergoing BCT [2–5]. IBTR is often classified as either true local recurrence (TR) or new ipsilateral primary tumor (NP). TR is described as ‘cases consistent with the regrowth of malignant cells not removed by surgery or not killed by radiotherapy’, whereas NP is described as ‘de novo cases of malignancies arising from mammary epithelial cells of the residual breast tissue’ [6]. The correct classification of IBTR status has significant implications in therapeutic decision-making and patient management; for example, TR patients will benefit from aggressive hormone therapy, chemotherapy, and/or additional radiotherapy, whereas NP patients may need only mild treatment. However, because of the inherent uncertainties of the clinical and pathologic criteria used for classification, the diagnostic tests to classify IBTR are subject to misclassification, and the validity of these tests has not been evaluated because there is no widely recognized classification standard for IBTR.

We thus developed a Bayesian method to (1) estimate the accuracy (i.e., sensitivity and specificity) of each diagnostic test to help clinicians make better decisions on which test to use on the basis of the sensitivity–specificity trade-off; (2) quantify the covariate effects on the probability of IBTR’s being NP

^aDivision of Biostatistics, University of Texas School of Public Health, 1200 Pressler St, Houston, Texas 77030, USA

^bDepartment of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA

^cDepartment of Biostatistics, The University of Texas MD Anderson Cancer Center, Texas 77030, USA

*Correspondence to: Sheng Luo, Division of Biostatistics, University of Texas School of Public Health, 1200 Pressler St, Houston, Texas 77030, USA.

†E-mail: sheng.t.luo@uth.tmc.edu

‡Supporting information may be found in the online version of this article.

§Sheng Luo and Min Yi contributed equally to this work.

and on the hazards of IBTR recurrence and death; and (3) provide useful prognostic information to future patients experiencing IBTR, that is, the prediction of the IBTR status and the median survival time (i.e., from IBTR to death).

1.1. A motivating dataset

Our methodology development has been motivated by the dataset consisting of 397 patients with invasive breast cancer who underwent BCT between 1970 and 2005 at The University of Texas MD Anderson Cancer Center and later developed IBTR as a first recurrence. The variables we collected included patient characteristics (age, race, family history of breast cancer, other cancer history), primary tumor characteristics (contralateral breast cancer, i.e., the occurrence of a second independent primary cancer in the other breast, location, histology, stage, size, estrogen receptor [ER] status), treatment characteristics (surgery, radiation), and patient status at last follow-up. ERs are tissue markers useful in assessing prognosis in breast cancer patients. An ER-positive (ER+) score indicates that estrogen is causing the tumor to grow and that the cancer should respond well to hormone-suppression treatments, whereas an ER-negative (ER-) score implies that the tumor is not driven by estrogen [7].

Ipsilateral breast tumor relapse patients in this dataset were classified as having either NP or TR by using two diagnostic tests based on readily available clinical and pathologic data. Test 1 is based on tumor location and histologic subtype: IBTR was defined as TR if the tumor was located within 3 cm of the primary tumor bed and its histologic subtype was consistent with that of the primary tumor; otherwise, IBTR was defined as NP [8–10]. Test 2 is based on tumor location, histologic subtype, and ER status: IBTR was defined as TR if the tumor was located within 3 cm of the primary tumor bed and its histologic subtype and ER status were consistent with those of the primary tumor; otherwise, IBTR was defined as NP [11, 12].

Because we use two criteria (tumor location and histology) in both tests 1 and 2 and because we use one additional criterion (ER status) in test 2 only, the following inherent relationships are true: NP by test 1 implies NP by test 2; TR by test 1 and NP by test 2 implies no change in tumor location and histology but change in ER status; and NP by test 1 and TR by test 2 is a null event. Therefore, test 2 always classifies more patients' IBTR as NP than does test 1; in our dataset, test 2 classified 213 IBTRs (53.6%) as NP, whereas test 1 classified 196 IBTRs (49.4%) as NP. If NP represents disease and TR represents nondisease in the conventional definitions of sensitivity and specificity, test 2 always has higher sensitivity but lower specificity than test 1. Although clinicians may use test 2 if higher sensitivity is preferred or use test 1 if higher specificity is preferred, it is essential to develop a methodological framework that provides accurate estimates of sensitivity and specificity for each diagnostic test to help clinicians select the test that best fits their preference.

1.2. Statistical challenges and solutions

Figure 1 displays the time plot of one patient, where t_1 represents the time from BCT to IBTR (referred to as time to relapse) and t_2 represents time from IBTR to cause-specific death or censoring (referred to as time from IBTR to death). An interesting feature of our patient dataset is the presence of two survival times and their strong correlation with IBTR status. To visualize the correlation, Figure 2 shows the Kaplan–Meier curves displaying differences in time to relapse (left panels) and time from IBTR to death (right panels) for patients classified as NP and TR by test 1 (top panels) and test 2 (bottom panels). Specifically, the left panels show that TR patients had shorter time to relapse than NP patients (mean: TR 5.4 years vs. NP 7.6 years; $p < 0.0001$ from test 1; mean: TR 5.4 years vs. NP 7.4 years; $p = 0.0003$ from test 2). The right panels of Figure 2 show that TR patients also had shorter time from IBTR to death than NP patients (mean, TR 5.5 years vs. NP 5.9 years; $p < 0.0001$ from test 1; mean, TR 5.6 years vs. NP 5.8 years; $p < 0.0001$ from test 2). If properly included in the model, we can use this survival information to help classify IBTR. Consider an extreme case as an example: If all NP patients lived



Figure 1. Time plot of one patient. t_1 is the time from BCT to IBTR, and t_2 is time from IBTR to cause-specific death or censoring. BCT, breast conservation therapy; IBTR, ipsilateral breast tumor relapse.

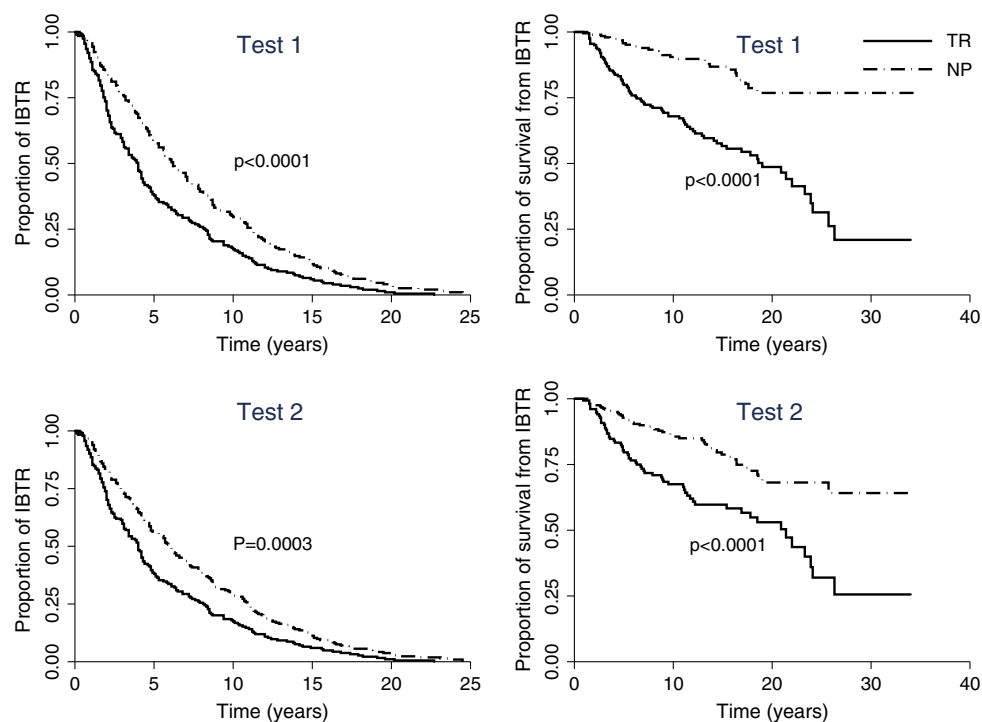


Figure 2. Kaplan–Meier curves displaying differences in time to relapse (left panels) and time from ipsilateral breast tumor relapse (IBTR) to death (right panels) for patients classified as having new primary tumors (NP) and true recurrences (TR) by test 1 (top panels) and test 2 (bottom panels).

longer than 15 years from IBTR and all TR patients died before 15 years from IBTR, then the length of survival time from IBTR would perfectly classify the IBTR status. Although our dataset does not reflect this perfect dichotomy, the large gaps between NP and TR patients' Kaplan–Meier curves provide useful information on the IBTR status.

To identify and quantify the risk factors associated with the binary IBTR status, researchers often use binomial regression models, which assume that the variables included in the model are accurately measured. In practice, categorical variables are often subject to misclassification, and continuous variables are subject to measurement error. Misclassification is due to many factors, including inaccuracy of data collection methods, limited sensitivity and specificity of the diagnostic tests, inadequacy of information derived from medical or other records, and recall bias in assessing exposure status [13]. Misclassification in binomial regression yields biased estimators of the association of covariates with response [14, 15].

Misclassification may occur in outcome variables, covariates, or both. In this article, we address misclassification in outcome variables. Magder and Hughes [16] showed that corrected odds ratios can be obtained by incorporating external estimates of test sensitivity and specificity into the likelihood for logistic regression. However, in many practical cases, a gold standard or a validation subsample either does not exist or may be too expensive or invasive to obtain. Therefore, it is impossible to have external estimates of sensitivity and specificity. In addition, Magder and Hughes [16] treated test sensitivity and specificity as fixed values and did not account for their uncertainty. To address this issue and to model the covariate effects in prevalence estimation, researchers have proposed various Bayesian models that use likelihood functions that are based on observed and latent variables [17–19] or observed variables only [20]. A unique feature of our proposed method is that the likelihood formulation incorporates both the test sensitivity and specificity and the correlated survival information, thus improving the estimation of the covariate effects on the probability of IBTR's being NP and the diagnostic test accuracy.

In the statistical literature of misclassification, a common assumption is conditional independence, that is, that multiple diagnostic tests are independent conditional on the true disease status. This assumption is relatively strong and unrealistic in many practical situations [21]. Some models have been proposed to relax this assumption, including a model with more than two latent classes [22–24], a random effects model [25], and a model with two additional parameters [26]. Because the two diagnostic tests used

in our dataset have special relationships, we propose a simple two-latent-classes model that naturally incorporates conditional dependence without using any additional parameters.

The purpose of this article is to develop a Bayesian method to model the binomial regression with misclassified binary outcome and the time to relapse and time from IBTR to death. We describe a Bayesian model, a method to relax conditional independence assumption, and model inference in Section 2. The proposed method is evaluated via extensive simulation in Section 3. We apply our method to analyze the breast cancer dataset in Section 4. Section 5 provides the discussion.

2. Statistical methodology

2.1. Model and notation

In this section, we consider two diagnostic tests for IBTR status and formulate the modeling framework. Let the subscript $j = 1, 2$ denote time to relapse and time from IBTR to death, respectively. Suppose y (1 if NP, 0 if TR) is the unobserved true IBTR status. Let y_1 and y_2 (1 if NP, 0 if TR) be the observed outcomes from two diagnostic tests. Let (p_k, q_k) , for $k = 1, 2$, be the sensitivity and specificity of test k given the true IBTR status y , that is, $p_k = p(y_k = 1 | y = 1)$ and $q_k = p(y_k = 0 | y = 0)$. We assume that p_k and q_k do not depend on covariates (non-differential assumption). Given covariate \mathbf{x} , we use binomial regression model for the probability that the IBTR being NP with $\pi(\mathbf{x}) = P(y = 1 | \mathbf{x}) = g^{-1}(\mathbf{x}\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a vector of regression coefficients and g^{-1} is the inverse of a link function (e.g., probit, logit, complementary log–log). Specifically, we use logit link function. The likelihood of observing outcomes y_1 and y_2 for one patient is as follows:

$$f(y_1, y_2 | \mathbf{x}) = [\pi(\mathbf{x})p_1p_2 + \bar{\pi}(\mathbf{x})\bar{q}_1\bar{q}_2]^{y_1y_2} [\pi(\mathbf{x})\bar{p}_1p_2 + \bar{\pi}(\mathbf{x})q_1\bar{q}_2]^{\bar{y}_1y_2} [\pi(\mathbf{x})p_1\bar{p}_2 + \bar{\pi}(\mathbf{x})\bar{q}_1q_2]^{y_1\bar{y}_2} [\pi(\mathbf{x})\bar{p}_1\bar{p}_2 + \bar{\pi}(\mathbf{x})q_1q_2]^{\bar{y}_1\bar{y}_2}, \quad (1)$$

where the overhead bar denotes 1 minus the variable (e.g., $\bar{\pi}(\mathbf{x}) = 1 - \pi(\mathbf{x})$). The likelihood derivation (Appendix A) assumes that two diagnostic tests are independent conditional on the true IBTR status (conditional independence assumption). This assumption will be relaxed in Section 2.2. The likelihood formulation involves only the binomial regression with misclassified outcome and is essentially identical to model (2) of Tu *et al.* [17] or model (2) of McInturff *et al.* [20].

Next, we extend the preceding model by including the time to relapse t_1 and the time from IBTR to death t_2 (Figure 1). Following the notation of Cox proportional hazards model, λ_{j0} is the baseline hazard function of survival time for t_j , with $j = 1, 2$. We consider two types of baseline hazard function: Weibull distribution (Appendix B) and piecewise constant function. Lawless and Zhan [27] and Feng *et al.* [28] illustrated that models using a piecewise constant baseline hazard yield good estimators for both fixed effects and frailty. Piecewise constant baseline hazard function has been widely used in the literature [29–34]. Given a set of fixed time points $0 = \tau_{j0} < \tau_{j1} < \dots < \tau_{jm} < \tau_{j(m+1)}$ and the baseline hazard vector $\mathbf{g} = (\mathbf{g}_1, \mathbf{g}_2)$ with $\mathbf{g}_j = (g_{j0}, g_{j1}, \dots, g_{jm})$, we define the piecewise constant hazard function as $\lambda_{j0}(t_j) = \sum_{l=0}^m g_{jl} I_{jl}(t_j)$, with $I_{jl}(t_j) = 1$ if $\tau_{jl} \leq t_j < \tau_{j(l+1)}$ and 0 otherwise. Let δ_j (1 if uncensored, 0 otherwise) be the censoring indicator for time t_j . Because the event for t_1 is IBTR and is observed in every patient, $\delta_1 \equiv 1$, whereas δ_2 can be either 1 or 0 because of censoring.

To account for the within-patient correlation between t_1 and t_2 , we use the shared frailty b and assume that b follows normal distribution with mean 0 and variance σ . We also assume that t_1 and t_2 are independent conditional on the frailty term b . The hazard function of survival time t_j is $\lambda_j^N(t_j) = \lambda_{j0}(t_j) \exp(\mathbf{x}\boldsymbol{\gamma}_j + b)$, where the superscript N denotes NP status. Note that the covariates included in the hazard function can differ or overlap with those in model (1). The corresponding survival function is $S_j^N(t_j) = \exp\left[-\exp(\mathbf{x}\boldsymbol{\gamma}_j + b) \int_0^{t_j} \sum_{l=0}^m g_{jl} I_{jl}(s) ds\right]$. If $\tau_{ja} \leq t_j < \tau_{j(a+1)}$, we can simplify the survival function to $S_j^N(t_j) = \exp\left\{-\exp(\mathbf{x}\boldsymbol{\gamma}_j + b) \left[\sum_{l=0}^{a-1} g_{jl}(\tau_{j(l+1)} - \tau_{jl}) + g_{ja}(t_j - \tau_{ja})\right]\right\}$. To solve for t_j ,

$$t_j^N = \tau_{ja} - \frac{\log S_j^N(t_j)}{g_{ja} \exp(\mathbf{x}\boldsymbol{\gamma}_j + b)} - \frac{1}{g_{ja}} \sum_{l=0}^{a-1} g_{jl}(\tau_{j(l+1)} - \tau_{jl}). \quad (2)$$

The condition $\tau_{ja} \leq t_j < \tau_{j(a+1)}$ imposes the following constraint: $-\exp(\mathbf{x}\boldsymbol{\gamma}_j + b) \sum_{l=0}^a g_{jl}(\tau_{j(l+1)} - \tau_{jl}) < \log S_j^N(t_j) \leq -\exp(\mathbf{x}\boldsymbol{\gamma}_j + b) \sum_{l=0}^{a-1} g_{jl}(\tau_{j(l+1)} - \tau_{jl})$. The hazard function for TR patients is

$\lambda_j^T(t_j) = \lambda_j^N(t_j)e^{\alpha_j} = \lambda_{j0}(t_j) \exp(\mathbf{x}\boldsymbol{\gamma}_j + b + \alpha_j)$, where α_j describes the additional hazard of being TR patients compared with NP patients, and the superscript T denotes TR status. Following the same procedure, we can derive $S_j^T(t_j)$, t_j^T , and the constraint of $S_j^T(t_j)$.

The likelihoods of observing t_j for NP and TR patients are $L_j^N = [\lambda_j^N(t_j)]^{\delta_j} S_j^N(t_j)$ and $L_j^T = [\lambda_j^T(t_j)]^{\delta_j} S_j^T(t_j)$, respectively. By multiplying the survival likelihood functions L_j^N and L_j^T into model (1), the observed likelihood conditional on frailty b for one patient is as follows:

$$L(\boldsymbol{\theta}|b) = [\pi(\mathbf{x})p_1p_2L_1^NL_2^N + \bar{\pi}(\mathbf{x})\bar{q}_1\bar{q}_2L_1^TL_2^T]^{y_1y_2} [\pi(\mathbf{x})\bar{p}_1p_2L_1^NL_2^N + \bar{\pi}(\mathbf{x})q_1q_2L_1^TL_2^T]^{\bar{y}_1\bar{y}_2} \cdot [\pi(\mathbf{x})p_1\bar{p}_2L_1^NL_2^N + \bar{\pi}(\mathbf{x})\bar{q}_1q_2L_1^TL_2^T]^{y_1\bar{y}_2} [\pi(\mathbf{x})\bar{p}_1\bar{p}_2L_1^NL_2^N + \bar{\pi}(\mathbf{x})q_1q_2L_1^TL_2^T]^{\bar{y}_1\bar{y}_2}, \quad (3)$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, p_1, q_1, p_2, q_2, \boldsymbol{\gamma}_1, \alpha_1, \boldsymbol{\gamma}_2, \alpha_2, \sigma, \mathbf{g})$ is the unknown parameter vector. The likelihood formulation (3) involves both the binomial regression with the misclassified outcome and the survival times. The marginal likelihood is $L(\boldsymbol{\theta}) = \int L(\boldsymbol{\theta}|b)f(b)db$. Because this integral cannot be evaluated analytically, we use Bayesian inference based on Markov chain Monte Carlo (MCMC) posterior simulation (Section 2.3).

2.2. Conditional dependence

As discussed in Section 1.1, the two diagnostic tests used to classify IBTR share some criteria (tumor location and histology). These two tests are likely to be highly correlated, even conditional on the true IBTR status. Therefore, the likelihood formulations (1) and (3) in Section 2.1, which are derived under conditional independence assumption, may not be valid. In this section, we discuss how to address this issue of conditional dependence.

Tumor location and histology are clinical and pathologic features, respectively, whereas ER status is measured by standard immunohistochemistry with antigen retrieval prior to antibody incubation. One can assume that these three features are different phenotypes representing the underlying disease trait (i.e., the true IBTR status), which contains all information that the three features share. Thus, we assume that ER status is independent of tumor location and histology (i.e., that ER is independent of the outcome of test 1), conditional on the true IBTR status. This assumption is much weaker than the conditional independence assumption in Section 2.1.

We define the indicator function I_E for ER status change (1 if ER status changes, 0 otherwise). Similar to the definitions of sensitivity and specificity, we define $p_E = P(I_E = 1|y = 1)$ and $q_E = P(I_E = 0|y = 0)$, respectively. Using the relationships between test 1 and test 2 illustrated in Section 1.1, the sensitivity and specificity of test 2 are as follows:

$$p_2 = P(y_1 = 1|y = 1) + P(y_1 = 0, I_E = 1|y = 1) = p_1 + \bar{p}_1p_E; q_2 = P(y_1 = 0, I_E = 0|y = 0) = q_1q_E. \quad (4)$$

Tests 1 and 2 are now dependent conditional on the true IBTR status.

The likelihood of observing outcomes y_1 and y_2 for one patient is

$$f(y_1, y_2|\mathbf{x}) = [\pi(\mathbf{x})p_1 + \bar{\pi}(\mathbf{x})\bar{q}_1]^{y_1y_2} [\pi(\mathbf{x})\bar{p}_1p_E + \bar{\pi}(\mathbf{x})q_1q_E]^{\bar{y}_1\bar{y}_2} [\pi(\mathbf{x})\bar{p}_1\bar{p}_E + \bar{\pi}(\mathbf{x})q_1q_E]^{\bar{y}_1\bar{y}_2}. \quad (5)$$

The derivation of the likelihood is detailed in Appendix C. After including the survival information, the observed likelihood conditional on frailty b for one patient becomes

$$L(\boldsymbol{\theta}|b) = [\pi(\mathbf{x})p_1L_1^NL_2^N + \bar{\pi}(\mathbf{x})\bar{q}_1L_1^TL_2^T]^{y_1y_2} [\pi(\mathbf{x})\bar{p}_1p_EL_1^NL_2^N + \bar{\pi}(\mathbf{x})q_1q_EL_1^TL_2^T]^{\bar{y}_1\bar{y}_2} \cdot [\pi(\mathbf{x})\bar{p}_1\bar{p}_EL_1^NL_2^N + \bar{\pi}(\mathbf{x})q_1q_EL_1^TL_2^T]^{\bar{y}_1\bar{y}_2}, \quad (6)$$

where unknown parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, p_1, q_1, p_E, q_E, \boldsymbol{\gamma}_1, \alpha_1, \boldsymbol{\gamma}_2, \alpha_2, \sigma, \mathbf{g})$. We can compute the estimates of p_2 and q_2 from the estimates of $\boldsymbol{\theta}$ using Equation (4). Throughout the article, we refer to models (1) and (5) as the ‘reduced model’ and models (3) and (6) as the ‘proposed model’ because of the fact that the reduced model is a special case of the proposed model when there is no survival information available.

2.3. Bayesian inference

In this section, we describe our Bayesian framework for parameter estimation. To obtain the parameter estimates, we use Bayesian inference based on MCMC posterior simulation. We use vague prior distributions for all parameters. Specifically, independent normal distribution with mean 0 and variance 10 is taken for each component of the parameter vectors β , γ_j , and α_j , where $j = 1, 2$. Assuming that the diagnostic tests are more accurate than the toss of a fair coin, we use uniform prior distribution $p_k, q_k \sim \text{Unif}(0.5, 1)$, for $k = 1, 2$, under the conditional independence case or use uniform prior distributions $p_1, q_1 \sim \text{Unif}(0.5, 1)$ and $p_E, q_E \sim \text{Unif}(0, 1)$ under the conditional dependence case. Whereas this accuracy assumption is slightly stronger than the identifiability condition, that is, $p+q > 1$, employed by Fujisawa and Izumi [35], we believe that the assumption is reasonable for tests in practical use. For the frailty variance σ and each component in baseline hazard vector g , we assume a noninformative inverse gamma prior distribution, $\text{IG}(0.01, 0.01)$, so that the distribution has mean of 1 and variance of 100. If we use a Weibull distribution as the baseline hazard function, the prior distribution for the scale and shape parameters is log-normal, that is, $\log(\lambda_j), \log(\nu_j)$ follows a normal distribution with mean 0 and variance 10. The MCMC sampler is implemented using WinBUGS software [36]. We use the trace plots available in WinBUGS and view the absence of apparent trend in the plots as evidence of convergence. In addition, we run multiple chains with disperse initial values and compute Gelman–Rubin scale reduction statistics \hat{R} to ensure \hat{R} of all parameters are smaller than 1.1.

To select between Weibull distribution and piecewise constant function as the baseline hazard function and to determine the covariates in the proposed method, we adopt a model selection approach using the deviance information criterion (DIC) proposed by Spiegelhalter *et al.* [37]. The DIC provides an assessment of model fit and a penalty for model complexity. The deviance statistics is defined as $D(\theta, b) = -2 \log f(y|\theta, b) + 2 \log h(y)$, where $f(y|\theta, b)$ is the likelihood function for the observed data vector y given the parameter vector θ and frailty b , and $h(y)$ denotes a standardizing function of the data alone that has no impact on model selection [38]. We define the DIC as $\text{DIC} = 2\bar{D} - D(\bar{\theta}, \bar{b}) = \bar{D} + p_D$, where $\bar{D} = E_{\theta, b|y}[D]$ is the posterior mean of the deviance, $D(\bar{\theta}, \bar{b}) = D(E_{\theta, b|y}[\theta, b])$ is the deviance evaluated at the posterior means $\bar{\theta}$ and \bar{b} , and $p_D = \bar{D} - D(\bar{\theta}, \bar{b})$ is the effective number of parameters. Smaller values of DIC indicate a better-fitting model. We use WinBUGS to compute DIC. This DIC definition applies the original definition of DIC to the conditional distribution, and it is the conditional DIC (i.e., DIC₇) defined in Celeus *et al.* [39].

3. Simulation studies

We evaluated the performance of our method using extensive simulation studies. We considered four covariates: $\mathbf{x} = (x_1, x_2, x_3, x_4)'$, where x_1 was continuous, generated from $N(47, 12^2)$ and centered at its rounded median; x_2, x_3 , and x_4 were binary and were generated from Bernoulli distributions with probabilities of 0.3, 0.1, and 0.3, respectively. We present the details of these variables in Section 4, and the numbers used to generate them resemble their characteristics. We simulated 100 datasets of 400 subjects each. To determine the burn-in iterations and assess the MCMC convergence and mixing properties, we examined the trace plots and the autocorrelations. We found that the chains converged reasonably fast and all achieved stationarity within 10,000 iterations with \hat{R} of all parameters smaller than 1.1. To ensure the convergence, we chose 25,000 iterations for burn-in for each MCMC sample, and we base the inference on the subsequent 25,000 iterations. The histories of eight parameters of interest from one randomly selected chain for one of the simulated datasets indicated reasonable convergence and mixing properties, even though, for clarity, only every 100th simulation was displayed (Web Figure 1). We noted similar good chain properties in all other examples presented in this article.

In the first set of simulation, we used the Weibull distribution as the baseline hazard function and simulated data structure consisting of an imperfectly classified binary outcome and the correlated survival information. We estimated the parameters of the proposed model using Weibull distribution as the baseline hazard function and examined the advantages of the proposed model over the logistic regression method and the reduced model under either conditional independence or conditional dependence assumption. We simulated the data using the following steps.

1. Simulated y using Bernoulli distribution with $\pi(\mathbf{x})$ generated from $\text{logit}[\pi(\mathbf{x})] = \mathbf{x}\beta$ with $\beta = (0.31, 0.006, -1.8, 0.7)'$ and $\mathbf{x} = (x_0, x_1, x_2, x_3)$.

2. Conditional on y , simulated the observed outcomes y_1 and y_2 with sensitivity and specificity $p_1 = 0.8$, $q_1 = 0.85$, $p_2 = 0.9$, and $q_2 = 0.7$ for the conditional independence case or $p_1 = 0.8$, $q_1 = 0.85$, $p_E = 0.15$, and $q_E = 0.75$ to get $p_2 = 0.830$ and $q_2 = 0.638$ for the conditional dependence case.
3. Simulated the shared frailty b from normal distribution $N(0, 0.1^2)$.
4. Simulated the survival probabilities S_j^N and S_j^T from uniform(0, 1) for $j = 1, 2$.
5. Conditional on y , generated the time to relapse t_1^N and t_1^T from Equation (2) with $\boldsymbol{\gamma}_1 = (-0.02, 0.2)'$, and the covariate vector being (x_1, x_4) , $\alpha_1 = 1$, and the Weibull distribution parameters $\lambda_1 = \nu_1 = 1$.
6. Conditional on y , generated the uncensored time from IBTR to death ζ^N and ζ^T from Equation (2) with $\boldsymbol{\gamma}_2 = (-0.004, 0.7)'$, the covariate vector being (x_1, x_4) , $\alpha_2 = 3$, and the Weibull distribution parameters $\lambda_2 = \nu_2 = 1$.
7. Simulated the censoring time from IBTR to death c_2 from uniform(0.1, 30). Let the observed time from IBTR to death $t_2^N = \min(\zeta^N, c_2)$ and $t_2^T = \min(\zeta^T, c_2)$ and obtained the censoring indicator δ_2 accordingly.
8. Repeated steps 1 to 7 until all subjects were generated.

The mean, standard error (the square root of the average of the posterior variances, denoted by SE), standard deviation (the standard deviation of the posterior means, denoted by SD), and coverage probabilities of 95% credible intervals estimated from logistic regression model (in which the second measurement y_2 was treated as the true y), the reduced model, and the proposed model are shown in Table I. When conditional independence was assumed, logistic regression gave severely biased parameter estimates, and the coverage probabilities were far away from the nominal level of 95% (Table Ia). The reduced model was valid in this scenario, giving consistent estimates and coverage probabilities close to 95% nominal level. Compared with the reduced model, the proposed model provided estimates with negligible bias and much smaller standard deviation for all parameters while retaining the coverage probabilities at 95%. The asymptotic relative efficiency (i.e., ARE, comparing the estimator based on the proposed model with the reduced model) in Table Ia indicates large efficiency gain in all parameter estimates (ranging from 5.378 to 1.690). These results indicated that the proposed model, which incorporates the additional correlated survival information, markedly improved the parameter estimation in terms of bias and efficiency.

Table Ib displays the estimation results under conditional dependence. Logistic regression had severe bias for all parameters. Because the reduced model correctly accounted for the conditional dependence structure, it was a valid model with consistent estimates and coverage probabilities close to 95%. In contrast, the proposed model's parameter estimates all had small bias and coverage probabilities that were reasonably close to 95%. These findings indicated that the conditional dependence between two diagnostic tests could be successfully addressed using the techniques in Section 2.2. The ARE in Table Ib indicated large efficiency gain in the proposed model because of the inclusion of the correlated survival time information.

Web Table I displays the simulation results when no misclassification was present ($p_j = q_j = 1$ for $j = 1, 2$) and conditional independence was assumed. All three methods provided comparable results, that is, the bias was negligible and the credible interval coverage probabilities were reasonably close to the nominal level of 95%. In addition, all estimates of sensitivity and specificity were close to 1. The results indicated that both the reduced and proposed models were robust in parameter estimation when no misclassification existed. The proposed model's estimates of the parameters ($\boldsymbol{\gamma}_1, \alpha_1, \boldsymbol{\gamma}_2, \alpha_2, \boldsymbol{g}, \sigma$) related to survival information had small bias, and the coverage probabilities of 95% credible intervals were close to 95% nominal levels (results not shown).

In the second set of simulation, we investigated the performance of our method when we used the piecewise constant function as the baseline hazard function. The data simulation followed the preceding steps with changes in step 5 using the baseline hazard vector $\boldsymbol{g}_1 = \boldsymbol{g}_2 = (0.05, 0.07, 0.1, 0.12, 0.16, 0.17, 0.18, 0.2, 0.15, 0.3, 0.1)$ at the fixed time points $\boldsymbol{\tau}_1 = (0, 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5, 25, 200)$ and $\boldsymbol{\tau}_2 = (0, 2, 4, 6, 8, 10, 12, 14, 20, 30, 40, 200)$, respectively. We simulated the datasets and estimated the parameters using piecewise constant baseline hazard function while assuming either conditional independence (Table IIa) or conditional dependence (Table IIc). Our model provided unbiased estimates, the SE close to the SD, and coverage probabilities close to 95%. In addition, we simulated the datasets using Weibull distribution as the baseline hazard

Table I. The estimates of the mean, standard error (SE), standard deviation (SD), and coverage probabilities (CP) of 95% credible intervals based on logistic regression (LR), the reduced model, and the proposed model assuming either conditional independence or conditional dependence.

		(a) Conditional independence							
	True pars	β_0	β_1	β_2	β_3	p_1	q_1	p_2	q_2
LR	Mean	0.599	0.004	-0.929	0.297				
	SD	0.142	0.009	0.245	0.375				
	SE	0.131	0.009	0.225	0.368				
	CP	0.400	0.930	0.030	0.740				
Reduced model	Mean	0.282	0.008	-1.932	0.635	0.812	0.849	0.894	0.690
	SD	0.222	0.012	0.501	0.436	0.064	0.042	0.037	0.052
	SE	0.327	0.013	0.537	0.478	0.072	0.052	0.048	0.066
	CP	0.990	0.950	0.950	0.980	0.980	0.990	0.990	0.980
Proposed model	Mean	0.288	0.007	-1.827	0.652	0.797	0.846	0.898	0.694
	SD	0.149	0.010	0.332	0.352	0.032	0.029	0.022	0.032
	SE	0.141	0.010	0.290	0.386	0.032	0.028	0.024	0.034
	CP	0.940	0.950	0.930	0.950	0.950	0.960	0.980	0.960
	ARE	5.378	1.690	3.429	1.533	5.062	3.449	4.000	3.768
		(b) Conditional dependence							
	True pars	β_0	β_1	β_2	β_3	p_1	q_1	p_2	q_2
LR	Mean	0.552	0.004	-0.769	0.219				
	SD	0.119	0.010	0.222	0.365				
	SE	0.130	0.009	0.224	0.363				
	CP	0.530	0.910	0.000	0.710				
Reduced model	Mean	0.330	0.008	-1.910	0.677	0.798	0.845	0.825	0.631
	SD	0.290	0.018	0.656	0.641	0.076	0.055	0.047	0.065
	SE	0.338	0.019	0.717	0.714	0.091	0.065	0.055	0.068
	CP	0.990	0.920	0.980	0.940	0.980	0.970	0.940	0.980
Proposed model	Mean	0.278	0.008	-1.834	0.661	0.803	0.846	0.838	0.633
	SD	0.175	0.012	0.330	0.389	0.039	0.034	0.035	0.036
	SE	0.196	0.012	0.346	0.439	0.042	0.039	0.038	0.040
	CP	0.960	0.950	0.960	0.990	0.990	0.980	0.970	0.970
	ARE	2.974	2.507	4.294	2.645	4.694	2.778	2.095	2.890

ARE is the asymptotic relative efficiency of the proposed model with respect to the reduced model.

function but estimated the parameters using piecewise constant function while assuming either conditional independence (Table IIb) or conditional dependence (Table IIc). Bias was small, SE was close to SD, and coverage probabilities were close to 95% in all parameters except α_1 and α_2 , whose estimates were slightly off the true values and whose coverage probabilities were off from the nominal value. These results indicated that our model could account for the misclassification, could recover the true parameters, and was generally robust to baseline hazard function misspecification.

4. Application to the breast cancer dataset

We applied the proposed method to the breast cancer patient dataset. We used three parallel chains with overdispersed initial values and ran each chain for 50,000 iterations. We discarded the first 25,000 iterations as burn-in; we based the parameter estimates on the remaining 25,000 iterations from each chain. We observed good mixing properties of the chains for the model parameter in the trace plots.

We considered four covariates in our method: x_1 represented age at breast cancer diagnosis (mean: 47.5 years; SD: 11.3 years); x_2 represented whether a distant recurrence developed in organs other

Table II. The estimates of the mean, standard error (SE), standard deviation (SD), and coverage probabilities (CP) of 95% credible intervals under various simulation scenarios.

True pars	Regression model				Time to recurrence				Time to death					
	β_0	β_1	β_2	β_3	p_1	q_1	p_2	q_2	γ_{11}	γ_{12}	α_1	γ_{21}	γ_{22}	α_2
	0.310	0.006	-1.800	0.700	0.800	0.850	0.900	0.700	-0.020	0.200	1.000	-0.004	0.700	3.000
(a) Simulate and estimate using piecewise constant function assuming conditional independence														
Mean	0.337	0.008	-1.849	0.674	0.794	0.850	0.898	0.700	-0.021	0.205	0.992	-0.004	0.692	2.948
SD	0.136	0.011	0.281	0.309	0.035	0.030	0.023	0.036	0.005	0.107	0.134	0.005	0.120	0.243
SE	0.149	0.010	0.294	0.426	0.032	0.029	0.025	0.035	0.005	0.122	0.134	0.005	0.141	0.245
CP	0.990	0.910	0.960	0.990	0.950	0.970	0.960	0.950	0.950	0.990	0.950	0.950	0.990	0.950
(b) Simulate from Weibull and estimate using piecewise constant function assuming conditional independence														
Mean	0.336	0.006	-1.808	0.693	0.797	0.856	0.896	0.702	-0.021	0.175	0.936	-0.004	0.680	2.746
SD	0.153	0.011	0.320	0.372	0.033	0.029	0.021	0.033	0.004	0.110	0.130	0.004	0.104	0.206
SE	0.147	0.010	0.293	0.426	0.032	0.028	0.025	0.035	0.005	0.116	0.125	0.005	0.129	0.222
CP	0.960	0.940	0.930	0.960	0.950	0.970	0.990	0.940	0.970	0.940	0.900	0.990	0.990	0.770
	0.310	0.006	-1.800	0.700	0.800	0.850	0.830	0.638	-0.020	0.200	1.000	-0.004	0.700	3.000
(c) Simulate and estimate using piecewise constant function assuming conditional dependence														
Mean	0.316	0.006	-1.813	0.668	0.803	0.857	0.836	0.642	-0.022	0.177	0.965	-0.005	0.699	2.984
SD	0.172	0.011	0.337	0.398	0.034	0.031	0.033	0.035	0.005	0.106	0.151	0.006	0.112	0.332
SE	0.165	0.011	0.308	0.449	0.034	0.033	0.031	0.037	0.005	0.124	0.141	0.006	0.148	0.278
CP	0.950	0.950	0.950	0.990	0.960	0.950	0.930	0.960	0.930	0.990	0.930	0.970	0.970	0.910
(d) Simulate from Weibull and estimate using piecewise constant function assuming conditional dependence														
Mean	0.348	0.008	-1.838	0.721	0.791	0.858	0.825	0.639	-0.019	0.173	0.946	-0.003	0.668	2.665
SD	0.139	0.012	0.297	0.398	0.032	0.033	0.030	0.038	0.004	0.106	0.133	0.004	0.097	0.239
SE	0.164	0.011	0.307	0.454	0.034	0.032	0.032	0.037	0.005	0.117	0.134	0.005	0.135	0.252
CP	0.990	0.920	0.970	0.970	0.960	0.960	0.960	0.920	0.960	0.950	0.910	0.970	0.990	0.740

than breasts (e.g., bones, lung, brain, liver; prevalence: 26.7%); x_3 represented whether contralateral breast carcinoma developed (prevalence: 12.8%); x_4 represented primary tumor stage ($x_4 = 1$ if more aggressive stage II or higher and 0 otherwise; prevalence: 28%) [40]. Using the quantiles of each survival time, we obtained the time points $\tau_1 = (0, 1.3, 2.2, 3.6, 4.8, 7.0, 9.7, 13.5, 24.6)$ and $\tau_2 = (0, 1.15, 1.9, 2.5, 3.7, 5.2, 7.7, 12.1, 31.4)$.

For model selection and comparison, we computed the DIC (illustrated in Section 2.3). Web Table 2 presents estimated DICs for some models using either the piecewise constant function or the Weibull distribution as the baseline hazard function. The piecewise constant model, with x_1 , x_2 , and x_4 in the binomial regression part and x_1 and x_4 in both survival parts of our model, had the lowest DIC value and was selected as the final model. To check the proportional hazard assumption in modeling the survival times, we plotted the product-limit estimates of log cumulative hazard rates for the covariates (e.g., tumor stage displayed in Web Figure 2). The absence of gross departure from the hypothesis of parallel curves suggested that the proportional hazard assumption was reasonable. In addition, to check the assumption of constant hazard ratio for TR patients compared with NP patients, we plotted in Web Figure 3 the product-limit estimates of log cumulative hazard rates of time to relapse (left panels) and time from IBTR to death (right panels) for patients classified as NP or TR by test 1 (top panels) and test 2 (bottom panels). The pattern of parallel curves indicated that the proportionality assumption for being NP was reasonable.

Table III provides the means, SDs, and 95% CIs from our model assuming conditional dependence as illustrated in Section 2.2. The rows labeled ‘Regression’ in Table III display the results of modeling the probability of having NP. A negative sign for a parameter β indicated a smaller probability of having NP or a larger probability of having TR. The odds ratio of the IBTR being NP for the patients who developed distant recurrence in organs other than breasts was 0.0011 (i.e., $\exp(-6.737)$; 95% CI: $[4.067e-5, 0.010]$) compared with the patients without distant recurrence. The covariates age at diagnosis and tumor stage were not statistically significant. These findings were consistent with the results of previous studies [8–10]. The sensitivity and specificity estimates of both tests were below 0.8, although test 2 had higher sensitivity but lower specificity than test 1. Because TR patients tend to have shorter survival time from IBTR to death and need more aggressive treatment than NP patients, the misclassification of TR patients into NP are likely to be more costly than vice versa. Thus, test 1 may be more preferable because of its higher specificity, and the proposed model provided the estimates of the test accuracy.

The rows labeled T_1 in Table III show the results of modeling the hazard of IBTR occurrence after BCT. The hazard rate of IBTR occurrence is 1.246 (95% CI: $[1.116, 1.377]$) for every 10-year increase in the age at breast cancer diagnosis. The relative risk of IBTR in patients with stage II or higher tumors relative to patients with stage I tumors was 1.415 (95% CI: $[1.052, 1.943]$). The TR status significantly

Table III. The means, standard deviations (SD), and 95% credible intervals (CI) of the parameters from the proposed model.

		Mean	SD	95% CI	
Regression	Intercept	1.929	1.291	-0.636	4.279
	Age at diagnosis	0.016	0.027	-0.032	0.072
	Distant recurrence	-6.737	1.425	-10.110	-4.655
	Tumor stage	-0.348	0.809	-1.736	1.421
	p_1	0.618	0.031	0.557	0.677
	q_1	0.769	0.042	0.684	0.847
	p_2	0.654	0.030	0.594	0.712
	q_2	0.700	0.044	0.610	0.784
T_1	Age at diagnosis	0.022	0.005	0.011	0.032
	Tumor stage	0.347	0.157	0.051	0.664
	α_1	0.682	0.161	0.367	1.004
T_2	Age at diagnosis	0.024	0.012	0.001	0.048
	Tumor stage	0.291	0.281	-0.263	0.850
	α_2	5.408	1.444	3.481	9.143
	σ	0.483	0.224	0.128	0.998

increased the risk of IBTR, with a relative risk of 1.978 (95% CI: [1.443, 2.729]) compared with NP patients. Finally, the rows labeled T_2 in Table III show the results of modeling the hazard of death. The hazard rate of death is 1.271 (95% CI: [1.010, 1.616]) for every 10-year increase in age at breast cancer diagnosis. Tumor stage was not associated with the risk of death, whereas TR status significantly increased the risk. The large risk differences between TR and NP statuses (indicated by the significant estimates of α_1 and α_2) are visually displayed in the large gaps in the Kaplan–Meier curves (Figure 2) and had been reported previously [8, 9, 41].

We further provided some insight into our model's ability to compute the subject-specific estimates and predictions for future patients who develop IBTR. We obtained data from 20 additional breast cancer patients who were treated at MD Anderson Cancer Center, had experienced IBTR, and were not included in the preceding analysis. We applied our model-fitting results to this new dataset. We first calculated the probability of the IBTR being NP and obtained a minimum probability of 0.01, mean probability of 0.40, and maximum probability of 0.98. Using 0.90 as cut point, we classified eight patients' IBTR as NP. On the basis of the IBTR status, we could compute each patient's median survival time. For example, one patient had breast cancer at age 50 years, IBTR at age 52 years, no distant recurrence, and a stage II or higher tumor. The estimated probability of her IBTR being NP was 0.02 and was therefore classified as TR. The estimated median survival time from IBTR to death was 1.4 years. In contrast, another patient had breast cancer at age 45 years, IBTR at age 54 years, had distant recurrence, and a stage I tumor. The estimated probability of her IBTR being NP was 0.95 and was therefore classified as NP. The estimated median survival time from IBTR was 14.3 years.

5. Discussion

In this article, we developed a Bayesian method to model binomial regression with the misclassified binary outcome and two correlated survival times (time to relapse and time from IBTR to death). We described a simple two-latent-classes model to relax the assumption of conditional independence between diagnostic tests conditional on the true IBTR status. Using extensive simulation, we found that our modeling framework corrects biases and provides more efficient estimates for the covariate effects on the probability of IBTR and the diagnostic test sensitivity and specificity compared with the method that does not use the survival information.

From the analysis of our patient dataset, we found that the sensitivity and specificity of both tests, which used only clinical and pathological criteria, were smaller than 0.8. A more accurate diagnostic test using the molecular criteria should be developed in the future, as pointed out by Huang *et al.* [8]. Our model will enable clinicians to make better-informed decisions about which diagnostic test to use and have better knowledge in identifying and quantifying the covariate effects on the probability of IBTR being NP and on the hazards. In addition, we can use the model-fitting results from the original dataset to provide new patients who develop IBTR the subject-specific prediction of some useful prognostic information such as the probability of the IBTR being NP and the median survival time from IBTR to death. This prognostic information is valuable for efficiently developing, targeting, and evaluating intervention, thus helping clinicians provide more suitable therapy for their patients, which in turn will improve patient outcomes. We can broadly apply our proposed method to many studies with a similar data structure consisting of imperfectly classified binary outcomes and survival information that is correlated with the outcomes. For example, tumor grade is an important prognostic indicator of pancreatic cancer survival after pancreatic resection [42–47], as higher tumor grade (III vs. I/II) is generally associated with poor survival [48]. Another example is that higher expression level of the chemokine CXCL12 (high vs. low/moderate) is an important predictor of poor survival in ovarian cancer [49]. However, the classification of tumor grade and CXCL12 expression level are subject to errors, and there is no gold standard available.

Adjustment for potential bias due to misclassification requires information on the misclassification structure to make the model identifiable [50]. The covariates in \mathbf{x} in modeling IBTR status are instrumental variables, and they make the reduced model identifiable when their number of different possible realizations is sufficient [51]. In addition, the survival information included in the proposed model is an important determinant of IBTR status. This is manifested by the clear dichotomy in the Kaplan–Meier curves displayed in Figure 2. As pointed out by one of the reviewers, when there are no correlated survival data and the probability of IBTR being NP is independent of the covariate, that is, $\pi(\mathbf{x}) \equiv \pi$, the proposed model reduces to the scenario of two conditionally independent tests and one population and hence lacks identifiability. We can find the detailed discussion of the identifiability, modeling, and

parameter estimation of multiple conditional dependent diagnostic tests in some recent literature, for example, Dendukuri and Joseph [52], Georgiadis *et al.* [53], and Jones *et al.* [54].

Bayesian inference using MCMC simulation described in this article produces reliable results. However, the model fitting was computationally intensive. For example, it took approximately 8 and 6 h for piecewise constant and Weibull baseline hazard function models, respectively, to get 50,000 samples for a single MCMC chain on a PC (Dell workstation Optiplex 960, Intel quad CPU at 3 GHz, 8 GB RAM). In contrast, it took only about 30 min and a few seconds, using the reduced model and the logistic regression method, respectively, on the same PC. Even though our implementation was slower, our proposed method improved the parameter estimation in bias and efficiency and produced subject-level predictions. It has been previously discussed that the improvement on parameter estimation is important and necessary [55].

Our modeling strategy had some limitations that we will address in our future research endeavors. One limitation was that we assumed both sensitivity and specificity to be non-differential, that is, that they did not depend on the covariates. In reality, there may be a subgroup of patients whose disease is less likely than that of other patients to be misclassified under both diagnostic tests. For example, most previously published studies assumed that the IBTR that occurred at the same location and with the same histology as the primary tumor was TR. Thus, the classification of patients with no change in tumor location and histology into the TR group is less prone to errors than the classification of other patients into the NP group.

Appendix A: Likelihood derivation of model (1)

The likelihood of observing outcomes y_1 and y_2 for one patient is

$$\begin{aligned}
 & f(y_1, y_2 | \mathbf{x}) \\
 &= p(y_1 = 1, y_2 = 1 | \mathbf{x})^{y_1 y_2} p(y_1 = 0, y_2 = 1 | \mathbf{x})^{\bar{y}_1 y_2} p(y_1 = 1, y_2 = 0 | \mathbf{x})^{y_1 \bar{y}_2} p(y_1 = 0, y_2 = 0 | \mathbf{x})^{\bar{y}_1 \bar{y}_2} \\
 &= [p(y_1 = 1, y_2 = 1 | y = 1, \mathbf{x}) p(y = 1 | \mathbf{x}) + p(y_1 = 1, y_2 = 1 | y = 0, \mathbf{x}) p(y = 0 | \mathbf{x})]^{y_1 y_2} \\
 &\cdot [p(y_1 = 0, y_2 = 1 | y = 1, \mathbf{x}) p(y = 1 | \mathbf{x}) + p(y_1 = 0, y_2 = 1 | y = 0, \mathbf{x}) p(y = 0 | \mathbf{x})]^{y_1 \bar{y}_2} \\
 &\cdot [p(y_1 = 1, y_2 = 0 | y = 1, \mathbf{x}) p(y = 1 | \mathbf{x}) + p(y_1 = 1, y_2 = 0 | y = 0, \mathbf{x}) p(y = 0 | \mathbf{x})]^{y_1 \bar{y}_2} \\
 &\cdot [p(y_1 = 0, y_2 = 0 | y = 1, \mathbf{x}) p(y = 1 | \mathbf{x}) + p(y_1 = 0, y_2 = 0 | y = 0, \mathbf{x}) p(y = 0 | \mathbf{x})]^{y_1 \bar{y}_2} \\
 &= [\pi(\mathbf{x}) p_1 p_2 + \bar{\pi}(\mathbf{x}) \bar{q}_1 \bar{q}_2]^{y_1 y_2} [\pi(\mathbf{x}) \bar{p}_1 p_2 + \bar{\pi}(\mathbf{x}) q_1 \bar{q}_2]^{y_1 \bar{y}_2} [\pi(\mathbf{x}) p_1 \bar{p}_2 + \bar{\pi}(\mathbf{x}) \bar{q}_1 q_2]^{y_1 \bar{y}_2} \\
 &[\pi(\mathbf{x}) \bar{p}_1 \bar{p}_2 + \bar{\pi}(\mathbf{x}) q_1 q_2]^{y_1 \bar{y}_2}, \tag{7}
 \end{aligned}$$

The second equality is derived from the law of total probability, and the last equality holds because we assume that two diagnostic tests are independent conditional on the true IBTR status (conditional independence assumption).

Appendix B: Formulation of Weibull distribution as baseline hazard function

We assume that baseline hazard function λ_{0j} for survival time t_j follows Weibull distribution $\lambda_{0j}(t_j) = \lambda_j t_j^{\nu_j}$, where λ_j is the scale parameter and ν_j is the shape parameter. The hazard function of survival time t_j for NP patients is $\lambda_j^N(t_j) = \lambda_{0j}(t_j) \exp(\mathbf{x} \boldsymbol{\gamma}_j + b)$, where the superscript N denotes NP status. The survival function is: $S_j^N(t_j) = \exp\left\{-\frac{\lambda_j}{\nu_j+1} t_j^{\nu_j+1} \exp(\mathbf{x} \boldsymbol{\gamma}_j + b)\right\}$. To solve for t_j , we have

$$\begin{aligned}
 & t_j^N = \left[-\frac{(\nu_j+1) \log S_j^N}{\lambda_j \exp(\mathbf{x} \boldsymbol{\gamma}_j + b)} \right]^{\frac{1}{\nu_j+1}}. \text{ The likelihood of observing } t_j \text{ is } L_j^N = \left[\lambda_j^N(t_j) \right]^{\delta_j} S_j^N(t_j). \text{ The hazard} \\
 & \text{function for TR patients is } \lambda_j^T(t_j) = \lambda_j^N(t_j) e^{\alpha_j} = \lambda_{j0}(t_j) \exp(\mathbf{x} \boldsymbol{\gamma}_j + b + \alpha_j), \text{ where } \alpha_j \text{ describes the} \\
 & \text{additional hazards of being TR patients compared with NP patients and the superscript T denotes TR sta-} \\
 & \text{tus. We can derive } S_j^T(t_j) = \exp\left\{-\frac{\lambda_j}{\nu_j+1} t_j^{\nu_j+1} \exp(\mathbf{x} \boldsymbol{\gamma}_j + \alpha_j + b)\right\}, t_j^T = \left[-\frac{(\nu_j+1) \log S_j^T}{\lambda_j \exp(\mathbf{x} \boldsymbol{\gamma}_j + \alpha_j + b)} \right]^{\frac{1}{\nu_j+1}}, \\
 & \text{and the likelihood } L_j^T = \left[\lambda_j^T(t_j) \right]^{\delta_j} S_j^T(t_j).
 \end{aligned}$$

Appendix C: Likelihood derivation of model (5)

Under conditional dependence assumption, we have

$$\begin{aligned} P(y_1 = 1, y_2 = 1) &= P(y_1 = 1) = \pi(\mathbf{x})p_1 + \bar{\pi}(\mathbf{x})\bar{q}_1 \\ P(y_1 = 0, y_2 = 1) &= P(y_1 = 0, I_E = 1) \\ &= P(y_1 = 0, I_E = 1|y = 1)P(y = 1) + P(y_1 = 0, I_E = 1|y = 0)P(y = 0) \\ &= P(y_1 = 0|y = 1)P(I_E = 1|y = 1)P(y = 1) + P(y_1 = 0|y = 0)P(I_E = 1|y = 0)P(y = 0) \\ &= \pi(\mathbf{x})\bar{p}_1 p_E + \bar{\pi}(\mathbf{x})q_1 \bar{q}_E \end{aligned}$$

$$\begin{aligned} P(y_1 = 1, y_2 = 0) &= 0 \\ P(y_1 = 0, y_2 = 0) &= P(y_1 = 0, I_E = 0) \\ &= P(y_1 = 0, I_E = 0|y = 1)P(y = 1) + P(y_1 = 0, I_E = 0|y = 0)P(y = 0) \\ &= P(y_1 = 0|y = 1)P(I_E = 0|y = 1)P(y = 1) + P(y_1 = 0|y = 0)P(I_E = 0|y = 0)P(y = 0) \\ &= \pi(\mathbf{x})\bar{p}_1 \bar{p}_E + \bar{\pi}(\mathbf{x})q_1 q_E. \end{aligned}$$

Note that $P(y_1 = 1, y_2 = 0)$ is a null event because of the inherent relationship between the two diagnostic tests explained in Section 1.2.

The likelihood of observing outcomes y_1 and y_2 for one patient is

$$\begin{aligned} f(y_1, y_2|\mathbf{x}) &= p(y_1 = 1, y_2 = 1|\mathbf{x})^{y_1 y_2} p(y_1 = 0, y_2 = 1|\mathbf{x})^{\bar{y}_1 y_2} p(y_1 = 1, y_2 = 0|\mathbf{x})^{y_1 \bar{y}_2} p(y_1 = 0, y_2 = 0|\mathbf{x})^{\bar{y}_1 \bar{y}_2} \\ &= [\pi(\mathbf{x})p_1 + \bar{\pi}(\mathbf{x})\bar{q}_1]^{y_1 y_2} [\pi(\mathbf{x})\bar{p}_1 p_E + \bar{\pi}(\mathbf{x})q_1 \bar{q}_E]^{\bar{y}_1 y_2} [\pi(\mathbf{x})\bar{p}_1 \bar{p}_E + \bar{\pi}(\mathbf{x})q_1 q_E]^{\bar{y}_1 \bar{y}_2}. \end{aligned} \quad (8)$$

Acknowledgements

The authors wish thank the editor, the associate editor, and the two referees for the helpful comments and suggestions that have led to an improvement in this article. Sheng Luo's research was partially supported by two NIH/NINDS grants U01 NS043127 and U01NS43128. The authors are grateful to our colleagues Drs. Wenyan Chan, Yong Chen, and Jing Ning for helpful discussion.

References

1. Fisher B, Anderson S, Bryant J, Margolese RG, Deutsch M, Fisher ER, Jeong JH, Wolmark N. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. *New England Journal of Medicine* 2002; **347**:1233–1241. DOI: 10.1056/NEJMoa022152.
2. Chen AM, Meric-Bernstam F, Hunt KK, Thames H, Oswald MJ, Outlaw ED, Strom EA, McNeese MD, Kuerer HM, Ross MI, Singletary SE, Ames FC, Feig BW, Sahin AA, Perkins GH, Schechter NR, Hortobagyi GN, Buchholz TA. Breast conservation after neoadjuvant chemotherapy: the MD Anderson Cancer Center experience. *Journal of Clinical Oncology* 2004; **22**:2303–2312. DOI: 10.1200/JCO.2004.09.062.
3. Fowle B, Solin LJ, Schultz DJ, Rubenstein J, Goodman RL. Breast recurrence following conservative surgery and radiation: patterns of failure, prognosis, and pathologic findings from mastectomy specimens with implications for treatment. *International Journal of Radiation Oncology* 1990; **19**:833–842. DOI: 10.1016/0360-3016(90)90002-2.
4. Rouzier R, Extra JM, Carton M, Falco MC, Vincent-Salomon A, Fourquet A, Pouillart P, Bournstyn E. Primary chemotherapy for operable breast cancer: incidence and prognostic significance of ipsilateral breast tumor recurrence after breast-conserving surgery. *Journal of Clinical Oncology* 2001; **19**:3828–3835.
5. Van Dongen JA, Voogd AC, Fentiman IS, Legrand C, Sylvester RJ, Tong D, van der Schueren E, Helle PA, van Zijl, K, Bartelink H. Long-term results of a randomized trial comparing breast-conserving therapy with mastectomy: European Organization for Research and Treatment of Cancer 10801 trial. *Journal of the National Cancer Institute* 2000; **92**:1143–1150. DOI: 10.1093/jnci/92.14.1143.
6. Veronesi U, Marubini E, Del Vecchio M, Manzari A, Andreola S, Greco M, Luini A, Merson M, Saccozzi R, Rilke F, Salvadori B. Local recurrences and distant metastases after conservative breast cancer treatments: partly independent events. *Journal of the National Cancer Institute* 1995; **87**:19–27. DOI: 10.1093/jnci/87.1.19.
7. Hunt KK, Robb GL, Storm EA, Ueno NT (eds). *Breast Cancer*. Springer: New York, 2001.
8. Huang E, Buchholz TA, Meric F, Krishnamurthy S, Mirza NQ, Ames FC, Feig BW, Kuerer HM, Ross MI, Singletary SE, McNeese MD, Strom EA, Hunt KK. Classifying local disease recurrences after breast conservation therapy based on location and histology: new primary tumors have more favorable outcomes than true local disease recurrences. *Cancer* 2002; **95**:2059–2067. DOI: 10.1002/cncr.10952.

9. Komoike Y, Akiyama F, Iino Y, Ikeda T, Tanaka-Akashi S, Ohsumi S, Kusama M, Sano M, Shin E, Suemasu K, Sonoo H, Taguchi T, Nishi T, Nishimura R, Haga S, Mise K, Kinoshita T, Murakami S, Yoshimoto M, Tsukuma H, Inaji H. Analysis of ipsilateral breast tumor recurrences after breast-conserving treatment based on the classification of true recurrences and new primary tumors. *Breast Cancer* 2005; **12**:104–111. DOI: 10.2325/jbcs.12.104.
10. Abd-Alla HM, Lotayef MM, Abou Bakr A, Moneer MM. Ipsilateral in-breast tumor relapse after breast conservation therapy: true recurrence versus new primary tumor. *Journal of the Egyptian National Cancer Institute* 2006; **18**:183–190.
11. Chaudary MA, Millis RR, Hoskins EO, Halder M, Bulbrook RD, Cuzick J, Hayward JL. Bilateral primary breast cancer: a prospective study of disease incidence. *British Journal of Surgery* 1984; **71**:711–714. DOI: 10.1002/bjs.1800710924.
12. Haffty BG, Carter D, Flynn SD, Fischer DB, Brash DE, Simons J, Ziegler AM, Fischer JJ. Local recurrence versus new primary: clinical analysis of 82 breast relapses and potential applications for genetic fingerprinting. *International Journal of Radiation Oncology* 1993; **27**:575–583.
13. Gordis L. *Epidemiology*, (4th edn). Saunders: Philadelphia, 2008.
14. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *American Journal of Epidemiology* 1977; **105**:488–495.
15. Neuhaus JM. Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika* 1999; **86**:843–855.
16. Magder LS, Hughes JP. Logistic regression when the outcome is measured with uncertainty. *American Journal of Epidemiology* 1997; **146**:195–203.
17. Tu XM, Kowalski J, Jia G. Bayesian analysis of prevalence with covariates using simulation-based techniques: applications to HIV screening. *Statistics in Medicine* 1999; **18**:3059–3073.
18. Paulino CD, Soares P, Neuhaus J. Binomial regression with misclassification. *Biometrics* 2003; **59**(3):670–675.
19. Branscum AJ, Johnson WO, Hanson TE, Gardner IA. Bayesian semiparametric ROC curve estimation and disease diagnosis. *Statistics in Medicine* 2008; **27**(13):2474–2496.
20. McInturff P, Johnson WO, Cowling D, Gardner IA. Modelling risk when binary outcomes are subject to error. *Statistics in Medicine* 2004; **23**(7):1095–1109.
21. Hui SL, Zhou XH. Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research* 1998; **7**:354–370.
22. Rindskopf D, Rindskopf W. The value of latent class analysis in medical diagnosis. *Statistics in Medicine* 1986; **5**:21–27.
23. Alvord WG, Drummond JE, Arthur LO, Biggar RJ, Goedert JJ, Levine PH, Murphy ELJ, Weiss SH, Blattner WA. A method for predicting individual HIV infection status in the absence of clinical information. *AIDS Research and Human Retroviruses* 1988; **4**(4):295–304.
24. Formann AK. Measurement errors in caries diagnosis: some further latent class models. *Biometrics* 1994; **50**:865–871.
25. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 1996; **52**:797–810.
26. Black MA, Craig BA. Estimating disease prevalence in the absence of a gold standard. *Statistics in Medicine* 2002; **21**(18):2653–2669.
27. Lawless JF, Zhan M. Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics* 1998; **26**(4):549–565.
28. Feng S, Wolfe RA, Port FK. Frailty survival model analysis of the national deceased donor kidney transplant dataset using Poisson variance structures. *Journal of the American Statistical Association* 2005; **100**(471):728–735.
29. Liu L, Yu Z. A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine* 2008; **27**(16):3105–3124.
30. Liu L, Ma JZ, O'Quigley J. Joint analysis of multi-level repeated measures data and survival: an application to the end stage renal disease (ESRD) data. *Statistics in Medicine* 2008; **27**(27):5679–5691.
31. Liu L, Huang X, O'Quigley J. Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* 2008; **64**(3):950–958.
32. Liu L, Huang X. The use of Gaussian quadrature for estimation in frailty proportional hazards models. *Statistics in Medicine* 2008; **27**(14):2665–2683.
33. Liu L. Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statistics in Medicine* 2009; **28**(6):972–986.
34. Liu L, Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2009; **58**(1):65–81.
35. Fujisawa H, Izumi S. Inference about misclassification probabilities from repeated binary responses. *Biometrics* 2000; **56**(3):706–711.
36. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; **10**(4):325–337.
37. Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 2002; **64**(4):583–639.
38. Carlin BP, Louis TA. *Bayesian Methods for Data Analysis*. Chapman & Hall/CRC: Boca Raton, 2009.
39. Celeux G, Forbes F, Robert C, Titterton DM. Deviance information criteria for missing data models. *Bayesian Analysis* 2006; **1**:651–674.
40. Singletary SE, Allred C, Ashley P, Bassett LW, Berry D, Bland KI, Borgen PI, Clark G, Edge SB, Hayes DF, Hughes LL, Hutter RV, Morrow M, Page DL, Recht A, Theriault RL, Thor A, Weaver DL, Wieand HS, Greene FL. Revision of the American Joint Committee on Cancer Staging System for Breast Cancer. *Journal of Clinical Oncology* 2002; **20**(17):3576–3577.
41. Smith TE, Lee D, Turner BC, Carter D, Haffty BG. True recurrence vs. new primary ipsilateral breast tumor relapse: an analysis of clinical and pathologic differences and their implications in natural history, prognoses, and therapeutic management. *International Journal of Radiation Oncology Biology Physics* 2000; **48**(5):1281–1289.

42. Geer RJ, Brennan MF. Prognostic indicators for survival after resection of pancreatic adenocarcinoma. *The American Journal of Surgery* 1993; **165**(1):68–73.
43. Lim JE, Chien MW, Earle CC. Prognostic factors following curative resection for pancreatic adenocarcinoma: a population-based, linked database analysis of 396 patients. *Annals of Surgery* 2003; **237**(1):74.
44. Winter JM, Cameron JL, Campbell KA, Arnold MA, Chang DC, Coleman JA, Hodgins MB, Sauter PK, Hruban RH, Riall TS, Schulick RD, Choti MA, Lillemoe KD, Yeo CJ. 1423 pancreaticoduodenectomies for pancreatic cancer: a single-institution experience. *Journal of Gastrointestinal Surgery* 2006; **10**(9):1199–1211.
45. Benassai G, Mastrorilli M, Quarto G, Cappiello A, Giani U, Forestieri P, Mazzeo F. Factors influencing survival after resection for ductal adenocarcinoma of the head of the pancreas. *Journal of Surgical Oncology* 2000; **73**(4):212–218.
46. Gebhardt C, Meyer W, Reichel M, Wünsch PH. Prognostic factors in the operative treatment of ductal pancreatic carcinoma. *Langenbeck's Archives of Surgery* 2000; **385**(1):14–20.
47. Kuhlmann KFD, de Castro SMM, Wesseling JG, ten Kate FJW, Offerhaus GJA, Busch ORC, Van Gulik TM, Obertop H, Gouma DJ. Surgical treatment of pancreatic adenocarcinoma: actual survival and prognostic factors in 343 patients. *European Journal of Cancer* 2004; **40**(4):549–558.
48. Robbins SL, Kumar V, Abbas AK, Cotran RS, Fausto N. *Robbins and Cotran Pathologic Basis of Disease*. WB Saunders Company: Philadelphia, 2010.
49. Popple A, Durrant LG, Spendlove I, Rolland P, Scott IV, Deen S, Ramage JM. The chemokine, cxcl12, is an independent predictor of poor survival in ovarian cancer. *British Journal of Cancer* 2012; **106**:1306–1313.
50. Ren D, Stone RA. A Bayesian adjustment for covariate misclassification with correlated binary outcome data. *Journal of Applied Statistics* 2007; **34**:1019–1034.
51. Nagelkerke NJ, Fidler V, Buwalda M. Instrumental variables in the evaluation of diagnostic test procedures when the true disease state is unknown. *Statistics in Medicine* 1988; **7**:739–744.
52. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001; **57**(1):158–167.
53. Georgiadis MP, Johnson WO, Gardner IA, Singh R. Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2003; **52**(1):63–76.
54. Jones G, Johnson WO, Hanson TE, Christensen R. Identifiability of models for multiple diagnostic testing in the absence of a gold standard. *Biometrics* 2010; **66**(3):855–863.
55. Yi M, Buchholz TA, Meric-Bernstam F, Bedrosian I, Hwang RF, Ross MI, Kuerer HM, Luo S, Gonzalez-Angulo AM, Buzdar AU, Symmans WF, Feig BW, Lucci A, Huang EH, Hunt KK. Classification of ipsilateral breast tumor recurrences after breast conservation therapy can predict patient prognosis and facilitate treatment planning. *Annals of Surgery* 2011; **253**:572–579.

Web-based Supporting Materials for “A Bayesian Model for Misclassified Binary Outcomes and Correlated Survival Data with Applications to Breast Cancer” by Sheng Luo, Min Yi, Xuelin Huang, and Kelly K. Hunt

Table 1: The estimates of the posterior mean (PM), standard error (SE), standard deviation (SD), and coverage probabilities (CP) of 95% credible intervals based on logistic regression (LR), the reduced model, and the proposed model when no misclassification is present and conditional independence is assumed.

	True pars	β_0	β_1	β_2	β_3	p_1	q_1	p_2	q_2
LR	PM	0.305	0.007	-1.797	0.605	—	—	—	—
	SD	0.131	0.009	0.270	0.381	—	—	—	—
	SE	0.127	0.009	0.265	0.380	—	—	—	—
	CP	0.940	0.950	0.990	0.930	—	—	—	—
Reduced model	PM	0.307	0.007	-1.822	0.618	0.995	0.995	0.995	0.995
	SD	0.132	0.009	0.276	0.390	0.005	0.005	0.005	0.005
	SE	0.125	0.009	0.266	0.358	0.005	0.005	0.005	0.005
	CP	0.940	0.940	0.970	0.910	—	—	—	—
Proposed model	PM	0.308	0.007	-1.824	0.618	0.995	0.995	0.995	0.995
	SD	0.132	0.009	0.276	0.390	0.005	0.005	0.005	0.005
	SE	0.126	0.009	0.265	0.359	0.005	0.005	0.005	0.005
	CP	0.950	0.950	0.960	0.900	—	—	—	—

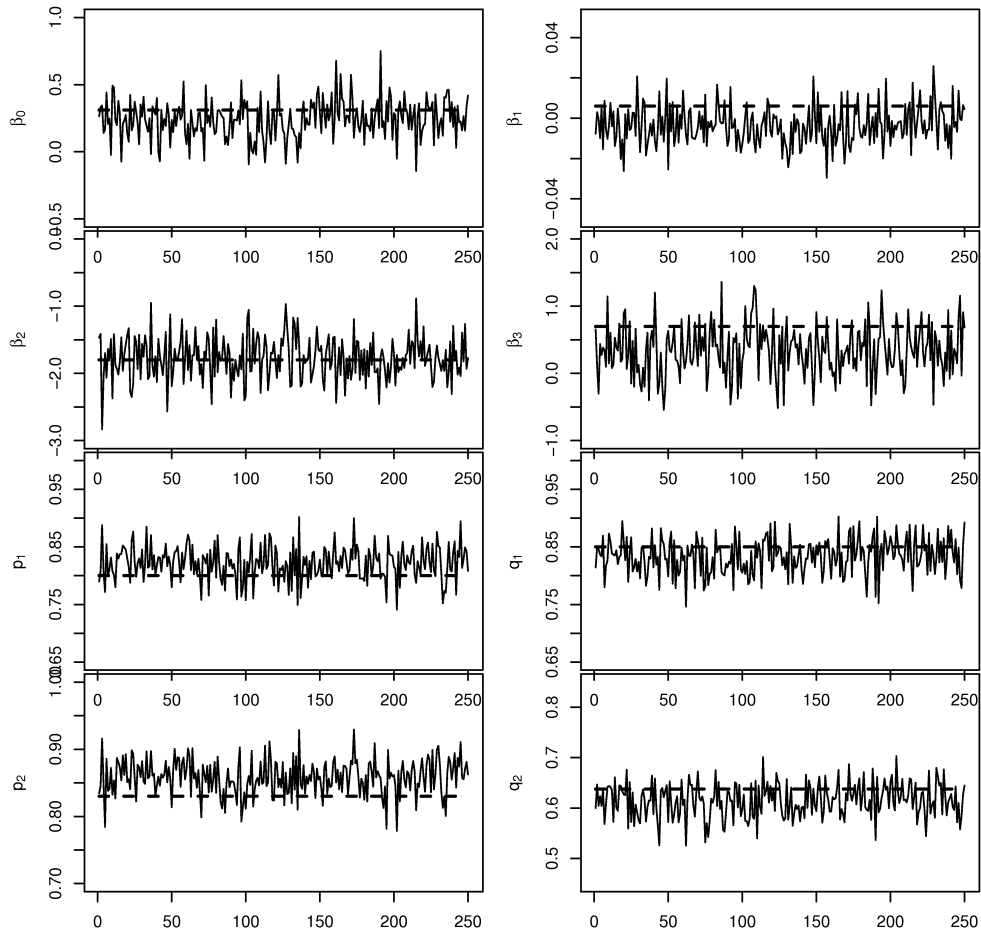


Figure 1: MCMC histories for 8 parameters of interest in one simulated dataset. The horizontal dash lines denote the true parameter values.

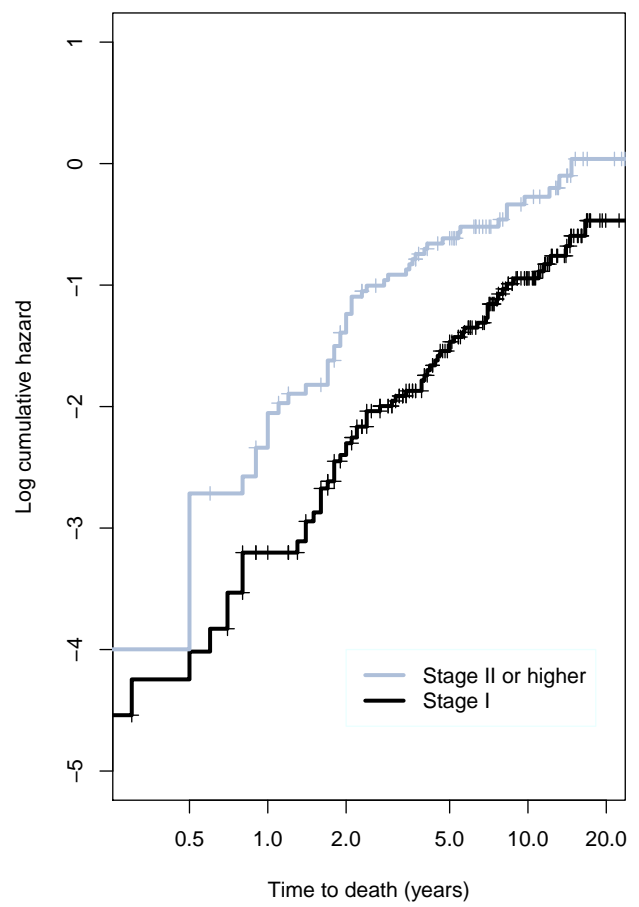
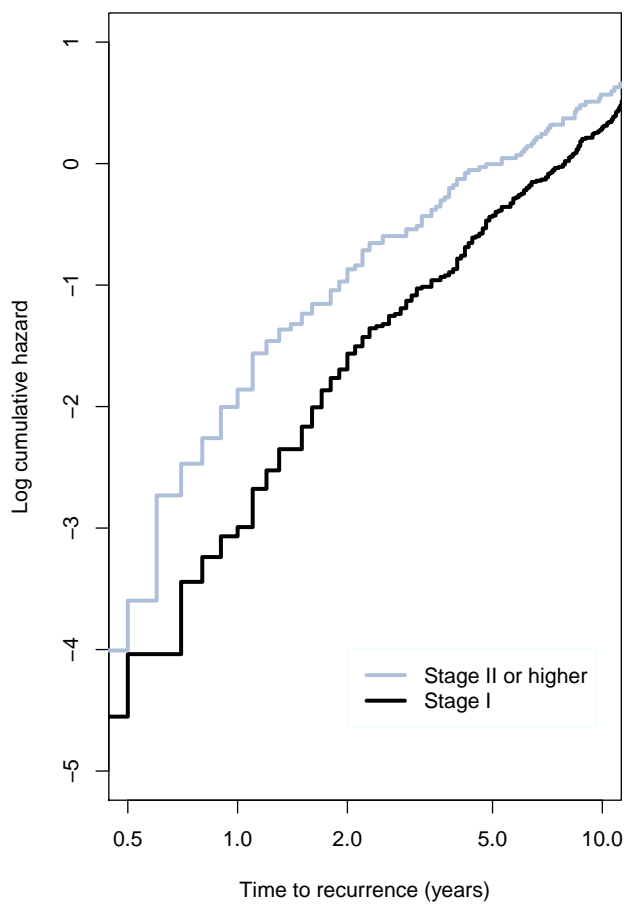


Figure 2: Log cumulative hazard curves for two tumor stage groups.

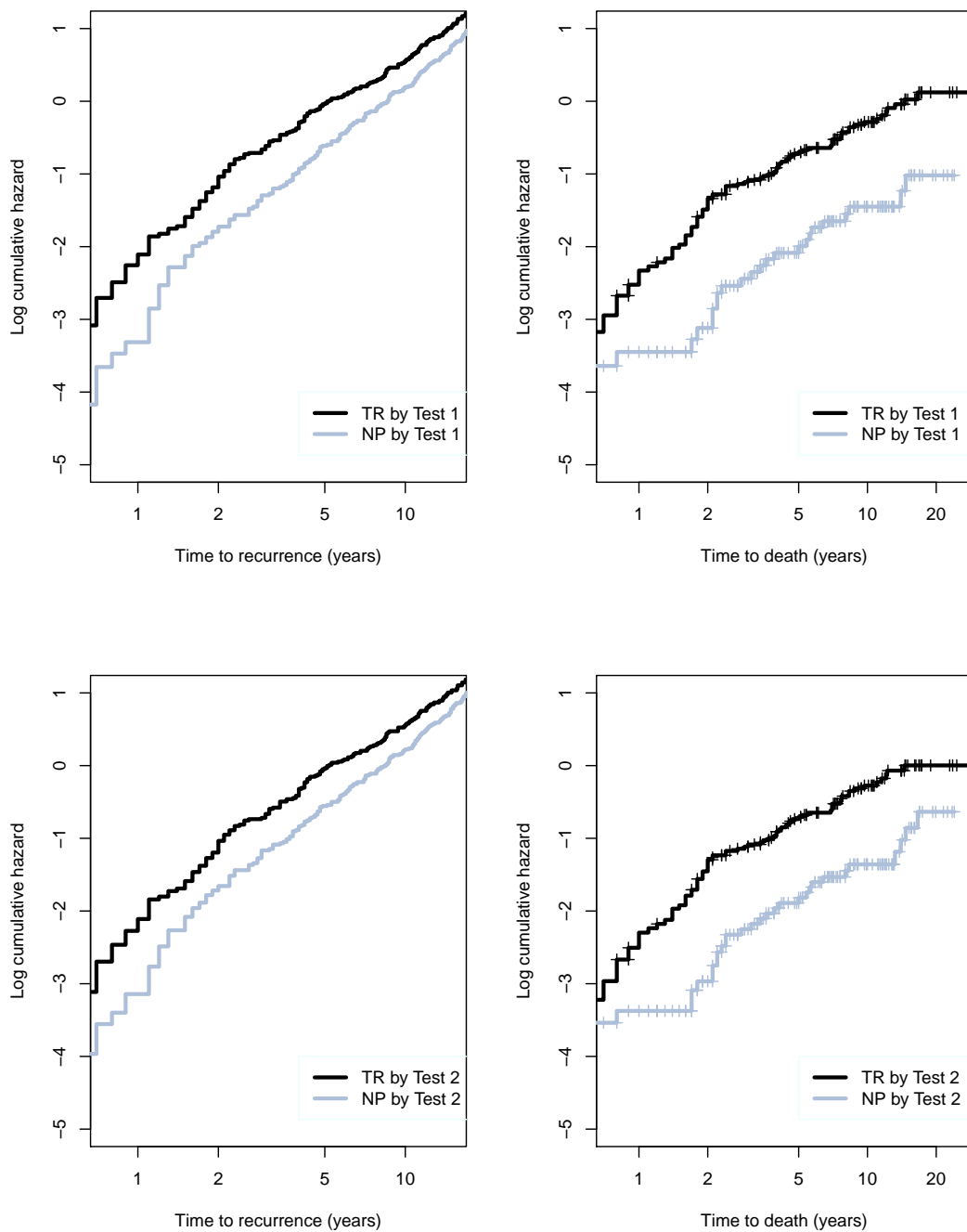


Figure 3: Log cumulative hazard curves of time to relapse (left panels) and time from IBTR to death (right panels) for patients classified as NP or TR by test 1 (top panels) and test 2 (bottom panels).

Table 2: Model comparison statistics for the MDACC dataset. The variables in the column labeled Model are in the binomial regression model with misclassification, while x_1 and x_4 are included in the survival models. DIC: deviance information criterion; \overline{D} : the posterior mean of the deviance; p_D : the effective number of parameters; x_1 : age at diagnosis; x_2 : distant recurrence; x_3 : contralateral; x_4 : tumor stage.

Baseline hazard	Model	DIC	\overline{D}	p_D
Piecewise constant	x_1, x_2	3493.9	3398.7	92.2
	x_1, x_3	3721.0	3671.0	50.0
	x_1, x_4	3740.0	3696.7	43.3
	x_1, x_2, x_3	3480.0	3377.9	102.5
	x_1, x_2, x_4^*	3458.4	3374.7	83.7
Weibull distribution	x_1, x_2, x_4	3500.5	3391.1	109.4

*Boldface indicates the preferred model.