

Joint model for a diagnostic test without a gold standard in the presence of a dependent terminal event

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

Breast cancer patients after breast conservation therapy often develop ipsilateral breast tumor relapse (IBTR), whose classification (true local recurrence versus new ipsilateral primary tumor) is subject to error, and there is no available gold standard. Some patients may die because of breast cancer before IBTR develops. Because this terminal event may be related to the individual patient's unobserved disease status and time to IBTR, the terminal mechanism is non-ignorable. This article presents a joint analysis framework to model the binomial regression with misclassified binary outcome and the correlated time to IBTR, subject to a dependent terminal event and in the absence of a gold standard. Shared random effects are used to link together two survival times. The proposed approach is evaluated by a simulation study and is applied to a breast cancer data set consisting of 4477 breast cancer patients. The proposed joint model can be conveniently fit using adaptive Gaussian quadrature tools implemented in SAS 9.3 (SAS Institute Inc., Cary, NC, USA) procedure NLMIXED. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: binomial regression; Cox model; frailty model; latent class model; informative censoring; tumor relapse

1. Introduction

The misclassification of a binary disease status in biomedical research due to imperfect sensitivity and specificity of a diagnostic test can have important ramifications for patient management. However, relative to a large body of methods on covariate measurement error, research on error-‘contaminated outcomes’, such as a misclassified diagnostic status, has been quite limited [1, 2]. When the interest is in predicting a binary misclassified disease status from a group of covariates, Neuhaus [3] and Neuhaus [4] have clearly demonstrated that a naive analysis that ignores the misclassification leads to biased results in a variety of settings. Most available methods require the use of a gold standard, which can provide estimates of the sensitivity and specificity of the imperfect measure, and then incorporate these external estimates into the likelihood to obtain corrected effect estimates [5–8]. Or if an internal validation subsample allowing comparison of an imperfect measure with the gold standard is available, some studies [9–12] have proposed a variety of techniques, for example, those based on likelihood or on weighted estimating equations. However, a gold standard or validation subsample may be unavailable, impractical [13, 14], expensive, time consuming, or unethical to perform on all subjects and is commonly difficult to obtain in clinical studies [15]. In the absence of a gold standard/validation subsample, few methods have been proposed for obtaining the accuracy of a single diagnostic test as most methods exploit the information contained in a battery of correlated imperfect diagnostic tests [16, 17]. For a single misclassified

^aDivision of Biostatistics, The University of Texas Health Science Center at Houston, 1200 Pressler St, Houston, TX 77030, U.S.A.

^bDepartment of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, U.S.A.

^cDepartment of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, U.S.A.

*Correspondence to: Sheng Luo, Division of Biostatistics, The University of Texas Health Science Center at Houston, 1200 Pressler St, Houston, TX 77030, U.S.A.

†E-mail: sheng.t.luo@uth.tmc.edu

binary endpoint, early research by Rindskopf and Rindskopf [18] and Formann [19] utilized a latent class analysis that allows for the estimation of the characteristics of indicators when an accurate diagnosis was unavailable. Duffy *et al* [20] presented assumption-dependent corrections to odds ratios when the disease was misclassified; however, this approach requires a priori knowledge of the relationship in order to apply a correction factor. Moreover, Joseph *et al* [21] used a Bayesian approach to estimating diagnostic error for a single diagnostic test by specifying priors for disease prevalence, sensitivity, and specificity, but determining which priors to use may not be straightforward in all applications.

In light of a single imperfect diagnostic test for disease status, adjustment for potential bias due to misclassification requires information on the misclassification structure [22]. Nagelkerke *et al* [23] suggested modeling the unobserved true disease classification as a logistic function of an instrumental variable, which is considered as an additional parameter serving to increase the outcome degrees of freedom. One can extend this instrumental variable framework whereby survival information is used to inform the disease classification. In oncology research, time to a survival event such as tumor relapse is likely informative for diagnosis, and its utilization in a model conditional on true disease status is conceptually analogous to conditioning on an instrumental variable. But before the disease status and the time to tumor relapse are observed, a terminal event such as disease-related death or dropout due to adverse event or severe adverse event could stop the follow-up of some patients. Because this terminal event may be related to the individual patient's unobserved disease status and time to tumor relapse, the terminal mechanism is non-ignorable. The dependent terminal event time is often termed 'dependent censoring' or 'informative censoring'. It has been shown in the joint modeling literature that ignoring the dependent censoring leads to biased estimates [24, 25]. However, to the best of our knowledge, there is no literature investigating the impact of ignoring the dependent censoring in the framework of binomial regression with misclassified binary outcome and a correlated survival time.

The goal of this article is to develop a joint analysis framework to model the binomial regression with misclassified binary outcome and the correlated time to ipsilateral breast tumor relapse (IBTR), subject to a dependent terminal event and in the absence of a gold standard. We applied the proposed model to a data set on breast cancer relapse diagnosis after breast conservation therapy (BCT). The article proceeds as follows. Section 2 describes a motivating data set. Section 3 formulates the joint statistical methodology. Section 4 presents results of a simulation study. Section 5 provides results of the motivating data set analysis and comparisons with previously published findings. Section 6 gives a discussion.

2. Motivating data

Tumor relapse after BCT for primary breast cancer has significant morbidity and mortality. Approximately 8–20% of patients undergoing BCT for primary breast cancer will experience IBTR, defined as the relapse of the tumor in the treated breast [26, 27]. It is important to accurately classify true local recurrence (TR) from a new ipsilateral primary tumor (NP), as treatment regimens are markedly different for the two conditions. The TR diagnosis is consistent with regrowth of malignant cells that are not removed by the initial surgery, while the NP diagnosis is consistent with a *de novo* case of malignancy arising from mammary epithelial cells of the residual breast tissue. The correct classification of IBTR status has significant implications in therapeutic decision-making and patient management; for example, patients experiencing a TR usually benefit from a more aggressive hormone therapy, chemotherapy, and/or additional radiotherapy, while NP patients often only require milder treatment. Currently, only imperfect histological test criteria are available to classify IBTR patients as TR or NP, for which there is no gold standard test. Statistically speaking, the binary diagnostic classification based on tumor histology suffers from misclassification.

The data from this study include 4477 patients with invasive breast cancer who underwent BCT between 1970 and 2010 at The University of Texas MD Anderson Cancer Center. A total of 397 patients later developed IBTR as a first relapse, and the remainder (4080) did not develop IBTR (censored for relapse process). The data have been described elsewhere in detail [17, 28–30] and will now be presented in the context of the current study. Relevant variables collected include patient characteristics (age, race, family history of breast cancer, and other cancer history), contralateral breast cancer (the occurrence of a second independent primary cancer in the other breast, location, histology, stage, size, estrogen receptor status), treatment characteristics (surgery, radiation), and patient status at last follow-up. We classified IBTR status as being either NP or TR using a single test based on tumor location and histologic subtype. Specifically, we defined IBTR 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

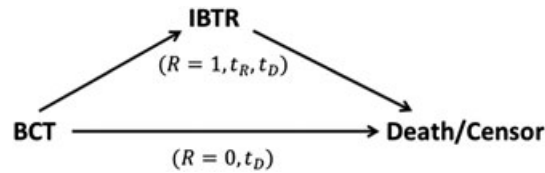


Figure 1. Data structures of patients with or without IBTR. R is an indicator of IBTR occurrence (1 if IBTR occurs, 0 otherwise), t_R is time from BCT to IBTR, and t_D is time from BCT to death or censoring. BCT, breast conservation therapy; IBTR, ipsilateral breast tumor relapse.

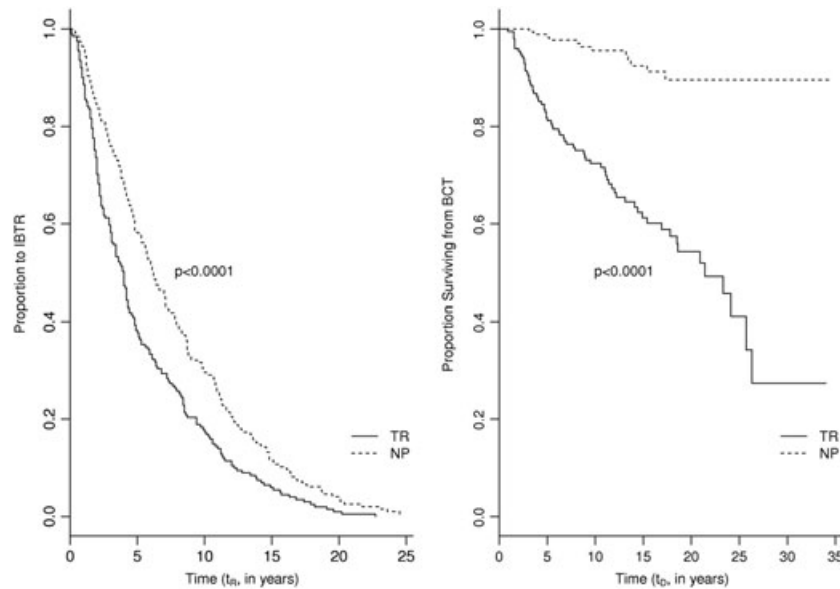


Figure 2. Kaplan–Meier curves for time from BCT to IBTR (left panel), time from BCT to death (right panel) in those who had IBTR ($R = 1$, sample size $N = 397$). p -values are from log rank tests. BCT, breast conservation therapy; IBTR, ipsilateral breast tumor relapse.

NP [28–30]. Among the 397 patients with IBTR, we classified 201 (50.6%) and 196 (49.4%) as TR and NP, respectively. However, because of the inherent uncertainties of the clinical and pathologic criteria used for classification, this diagnostic test to classify IBTR was subject to misclassification.

Figure 1 displays the data structures of the patients with or without IBTR, where R is an indicator of IBTR occurrence (1 if IBTR occurs, 0 otherwise), t_R is time from BCT to IBTR (referred to as time to IBTR), and t_D is time from BCT to breast cancer related death or censoring (referred to as time to death). While all patients have observed t_D , the patients with IBTR ($R = 1$) have one additional observable time t_R . Two interesting features of this patient data set are as follows: (i) the IBTR status is strongly correlated with both time to IBTR and time to death for the patients with IBTR; and (ii) the occurrence of IBTR is strongly correlated with time to death for all patients. To visualize the first correlation, Figure 2 displays the Kaplan–Meier curves showing differences in time to IBTR (t_R , left panels) and time to death (t_D , right panels) for patients with IBTR ($R = 1$, sample size $N = 397$) and classified as either NP or TR. The left panel indicates that TR patients have shorter time to IBTR than NP patients (log rank test p -value < 0.0001), while the right panel indicates that TR patients have shorter time to death than NP patients (log rank test p -value < 0.0001).

Among a total of 4477 patients, 251 died of breast cancer. The occurrence of IBTR is likely informative for the time to breast cancer death. To visualize this correlation, Figure 3 displays the Kaplan–Meier curves showing differences in time to death for patients with or without IBTR. The patients with IBTR have shorter time to death than the ones without IBTR (log rank test $p < 0.0001$). Hence, the breast cancer-related death is likely to be dependent terminal event. The purpose of this article is to develop a joint model of a misclassified binary outcome and time to relapse with dependent terminal event in order to (i) estimate the sensitivity and specificity of the diagnostic test and (ii) quantify covariate effects on the probability of IBTR being NP and on the hazards of IBTR and the terminal event.

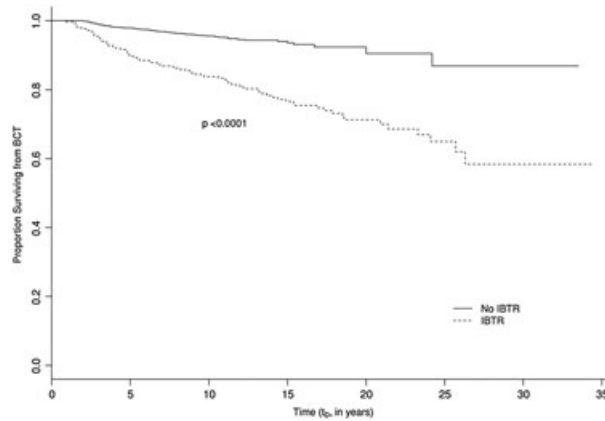


Figure 3. Kaplan–Meier curves for time from BCT to death of all patients, sample size $N = 4477$, with associated log rank test p -value. BCT, breast conservation therapy.

3. Statistical models and likelihood inference

3.1. Model and notation

In this section, we formulated a joint modeling framework for a single diagnostic test for IBTR status. Suppose R is an indicator of IBTR occurrence (1 if IBTR occurs, 0 otherwise). Let t_R be the time to IBTR and $t_D = \min(C, D)$, the minimum of an independent censoring time C and a dependent terminal event time D (e.g., breast cancer-related death), respectively, both measured from the BCT. Let t_D denote time to death. Let δ (1 if breast cancer death is observed, 0 otherwise) be the censoring indicator. Let $t = \min(t_R, t_D)$. Thus, $t = t_R$ and $R = 1$ for the patients with IBTR, and $t = t_D$ and $R = 0$ for the patients without IBTR. Suppose y (1 if NP, 0 if TR) is the unobserved true IBTR status. Let z (1 if NP, 0 if TR) be the observed outcome from the error-prone diagnostic test. Let p and q be the sensitivity and specificity of the test given the true IBTR status y , that is, $p = P(z = 1|y = 1)$ and $q = P(z = 0|y = 0)$. Hence, the sensitivity and specificity are the probabilities of correctly classify NP and TR patients, respectively.

Denote \mathbf{x}_P , \mathbf{x}_R , and \mathbf{x}_D as covariate vectors for the probability of IBTR being NP, time to IBTR, and time to death, respectively. Three covariate vectors can be the same or different. We made the following assumptions: (i) the sensitivity p and specificity q do not depend on covariates (non-differential assumption); (ii) the times to IBTR and breast cancer death for the same patient are correlated through a shared frailty $u \sim N(0, \sigma_u^2)$; and (iii) the diagnostic test results, time to IBTR, and time to death are independent conditional on the frailty u and covariate vectors. The binomial regression model for the probability that the IBTR being NP is

$$\pi(\mathbf{x}_P) = P(y = 1|\mathbf{x}_P) = g^{-1}(\mathbf{x}_P \boldsymbol{\alpha}) \quad (1)$$

where $\boldsymbol{\alpha}$ is the corresponding vector of regression coefficients and g^{-1} is the inverse of a link function (e.g., probit, logit, and complementary log-log). Specifically, we used the logit link function in this article. The likelihood of observing outcome z for one patient is $f(z|\mathbf{x}_P) = [\pi p + \bar{\pi} q]^z [\pi \bar{p} + \bar{\pi} q]^{1-z}$, where the overhead bar denotes 1 minus the variable (e.g., $\bar{\pi} = 1 - \pi$).

The hazard of breast cancer death is modeled by

$$\lambda(t_D) = \lambda_0(t_D) \exp(\mathbf{x}_D \boldsymbol{\eta} + u) \quad (2)$$

where $\lambda_0(t_D)$ is the baseline hazard, with corresponding survival function being $S(t_D) = \exp\left[-\exp(\mathbf{x}_D \boldsymbol{\eta} + u) \int_0^{t_D} \lambda_0(t) dt\right]$. The hazard function of time to IBTR t_R is

$$r^N(t) = r_0(t) \exp(\mathbf{x}_R \boldsymbol{\beta} + \gamma u) \quad (3)$$

where the superscript N denotes NP status, $r_0(t)$ is baseline hazard for IBTR. The corresponding survival function $S^N(t) = \exp\left[-\exp(x_R\beta + \gamma u) \int_0^t r_0(x)dx\right]$. The hazard for TR patients is $r^T(t) = r_0(t) \exp(x_R\beta + \zeta + \gamma u)$, where ζ describes the additional hazard of being TR patients compared with NP patients and the superscript T denotes TR status. We linked together the hazard functions of IBTR and death via the shared random effect u , and the parameter γ measures the association between two models. The likelihood of the relapse process is

$$L_1(\theta|u) = \left\{ [\pi p r^N(t) S^N(t) + \bar{\pi} \bar{q} r^T(t) S^T(t)]^z [\pi \bar{p} r^N(t) S^N(t) + \bar{\pi} q r^T(t) S^T(t)]^{\bar{z}} \right\}^R \cdot \left\{ [\pi S^N(t) + \bar{\pi} S^T(t)] \right\}^{1-R} \quad (4)$$

where the unknown parameter vector $\theta = (\alpha, \beta, \eta, \sigma_u, \zeta, \gamma, p, q)$. The likelihood of the death process is

$$L_2(\theta|u) = [\lambda(t_D)]^\delta S(t_D) \quad (5)$$

Thus, the marginal likelihood for one patient is

$$L(\theta) = \int L(\theta|u) f(u) du = \int L_1(\theta|u) L_2(\theta|u) f(u) du \quad (6)$$

We referred the proposed model, $L(\theta)$, to as joint model. In addition, the reduced model refers to the analysis of maximizing $L_1(\theta|u)$ and $L_2(\theta|u)$ separately by assuming that the time to IBTR is independent of time to death (i.e., $\gamma = 0$).

3.2. Maximum likelihood estimation

The marginal likelihood function in (6) involves an integral with respect to random effects u , and we cannot evaluate this integral analytically. We can use numerical integration such as Laplace approximation [31] and Gaussian quadrature [32] for estimation. Liu *et al* [33] pointed out that both methods generally perform well, with the Laplace approximation being much faster in high dimensional random effects settings. However, Laplace approximation is not yet available in commercial software packages for fitting nonlinear mixed effects models, while Gaussian quadrature can be conveniently implemented in SAS procedure NLMIXED. Thus, we adopted adaptive Gaussian quadrature to approximate the integral in model (6). In numerical analysis, a quadrature rule is an approximation of the definite integral by a weighted sum of function values at specified points within the domain of integration. Moreover, the adaptive Gaussian quadrature method accounts for the shape of the likelihood when placing quadrature points and provides a better approximation than the non-adaptive Gaussian quadrature with equally spaced points [34]. In addition to accurate parameter estimates and available standard error (SE) estimates, this estimation approach possesses the advantage of easy implementation because SAS procedure NLMIXED requires inputting the likelihood (conditional on random effects) explicitly and the approximation of the marginal likelihood can be directly maximized.

However, with the unspecified baseline hazard functions $r_0(t)$ and $\lambda_0(t_D)$, there are non-parametric terms (i.e., baseline hazard functions) in the likelihood functions $L_1(\theta|u)$ in (4) and $L_2(\theta|u)$ in (5). Hence, the adaptive Gaussian quadrature method is not directly applicable. Instead, we used a piecewise constant function to approximate the baseline hazard function $r_0(t)$ and $\lambda_0(t_D)$. It has been shown that survival models with a piecewise constant baseline hazard with 8 to 10 intervals yield good estimators for both fixed effects and frailty [35, 36], although fixed cut points need to be specified a priori. It is more flexible than the a priori choices of parametric baseline hazard distribution (e.g., Weibull distribution), and it retains enough model structure [32]. In literature, studies have widely used [32, 37, 38] piecewise constant baseline hazard function. Given a set of fixed time points $\tau_D = (\tau_{D0}, \dots, \tau_{Dm})$ with the constraint $0 = \tau_{D0} < \tau_{D1} < \dots < \tau_{Dm}$, and the baseline hazard vector $g_D = (g_{D0}, g_{D1}, \dots, g_{Dm-1})$ for the time to death t_D , we defined the piecewise constant baseline hazard function as $\lambda_0(t_D) = \sum_{l=0}^{m-1} g_{Dl} I_l(t_D)$, with indicator function $I_l(t_D) = 1$ if $\tau_l \leq t_D < \tau_{l+1}$ and 0 otherwise. Similarly, we can define the piecewise constant baseline hazard function $r_0(t)$ for time to IBTR on the basis of a fixed time points vector τ_R and a baseline hazard vector g_R . We can approximate the marginal likelihood $L(\theta)$ in model (6) by replacing $r_0(t)$ and $\lambda_0(t_D)$ by their piecewise constant counterparts. The resulting parametric likelihood is maximized by SAS procedure NLMIXED.

We presented an example of SAS code for fitting the proposed model with piecewise constant baseline hazard functions in the Appendix.

4. Simulation studies

This section presents results from a simulation study of two settings to compare the performance of the proposed joint model and reduced model. The simulated data structure is similar to but simpler than the motivating data set. In each setting, we generated 3000 data sets with sample size $N = 500$ using the following algorithm.

1. Simulate y from a Bernoulli distribution with π generated from $\text{logit}(\pi) = \mathbf{x}_P \boldsymbol{\alpha}$ in model (1) with $\boldsymbol{\alpha} = (-2, 3, 4)$ and $\mathbf{x}_P = (1, x_1, x_2)$ with $x_1 \sim N(0, 1)$ and $x_2 \sim \text{Bernoulli}(0.3)$.
2. Simulate the frailty $u \sim N(0, \sigma_u^2)$ with $\sigma_u^2 = 3$.
3. Set the baseline hazard function for the time to death as $\lambda_0(t) = t^{-0.5}$ in model (2) and set $\boldsymbol{\eta} = (0.5, 0.5)$ and $\mathbf{x}_D = (x_1, x_2)$. Simulate S_D from $\text{uniform}(0, 1)$ and generate uncensored time D for the terminal event, subject to independent censoring time C . Obtain $t_D = \min(C, D)$ and the censoring indicator δ .
4. Set the baseline hazard function for the time to relapse as $r_0(t) = 1.5t^{-0.5}$ in model (3) and set $\boldsymbol{\beta} = (0.5, 0.5)$, $\mathbf{x}_D = (x_1, x_2)$, and $\zeta = 3$. Set $\gamma = 0$ for setting I and set $\gamma = 0.5$ for setting II. Conditional on y , simulate S^N and S^T from $\text{uniform}(0, 1)$ and generate t_R for the relapse process. If $t_R \leq t_D$, set $R = 1$ and simulate z with sensitivity p and specificity q . If $t_R > t_D$, set $R = 0$ and IBTR is not observed.
5. Repeat steps 1 to 4 until all patients are generated.

In setting I, the terminal event is independent of the relapse process subject to misclassification (i.e., $\gamma = 0$), and the reduced model is the true model. In setting II, the terminal event is dependent on the relapse process subject to misclassification (i.e., $\gamma = 0.5$), which is likely a reasonable estimate for the application of interest, considering Figure 3. So in setting II, the proposed joint model is the true model. The simulation results presented in Tables I and II report the bias, SE of parameter estimates, mean of SE estimates (SEM), and coverage probability (CP) of 95% confidence interval (CI) for each parameter of interest under the joint and reduced models defined in Section 3.

Table I suggests that setting I with no correlation (i.e., $\gamma = 0$), the reduced model gives reasonable parameter estimates; that is, the bias is negligible, SE is close to SEM, and the CI coverage probabilities are reasonably close to 0.95 nominal level. In comparison, the joint model generates comparable results

Table I. Simulation results of fitting the proposed joint model and the reduced model for setting I in which the terminal event is independent on the relapse process subject to misclassification.

Parameter	Joint model				Reduced model			
	Bias	SE	SEM	CP	Bias	SE	SEM	CP
For the probability of IBTR being NP								
$\alpha_0 = -2$	-0.082	0.319	0.312	0.952	-0.056	0.316	0.306	0.952
$\alpha_1 = 3$	0.102	0.433	0.419	0.947	0.083	0.427	0.414	0.946
$\alpha_2 = 4$	0.117	0.624	0.608	0.947	0.094	0.624	0.603	0.946
$p = 0.8$	0.006	0.056	0.055	0.921	0.002	0.056	0.055	0.933
$q = 0.9$	-0.000	0.020	0.020	0.940	0.000	0.021	0.020	0.939
For the time to IBTR								
$\beta_1 = 0.5$	0.014	0.091	0.091	0.946	0.003	0.088	0.085	0.938
$\beta_2 = 0.5$	0.012	0.185	0.183	0.950	0.003	0.181	0.175	0.941
$\zeta = 3$	0.108	0.266	0.278	0.953	0.031	0.238	0.233	0.945
For the time to terminal event								
$\eta_1 = 0.5$	0.075	0.128	0.121	0.950	0.004	0.071	0.069	0.944
$\eta_2 = 0.5$	0.076	0.188	0.190	0.959	0.002	0.139	0.138	0.950
$\gamma = 0$	0.079	1.044	1.863	0.998				

SE, standard error; SEM, mean of the standard error estimates; CP, coverage probability; IBTR, ipsilateral breast tumor relapse; NP, new ipsilateral primary tumor.

Table II. Simulation results of fitting the proposed joint model and the reduced model for setting II in which the terminal event is dependent on the relapse process subject to misclassification.

Parameter	Joint model				Reduced model			
	Bias	SE	SEM	CP	Bias	SE	SEM	CP
For the probability of IBTR being NP								
$\alpha_0 = -2$	-0.015	0.355	0.338	0.915	0.061	0.345	0.358	0.926
$\alpha_1 = 3$	0.047	0.459	0.461	0.926	-0.097	0.459	0.476	0.915
$\alpha_2 = 4$	0.036	0.644	0.660	0.940	-0.118	0.651	0.688	0.935
$p = 0.8$	0.009	0.072	0.069	0.916	-0.008	0.073	0.074	0.942
$q = 0.9$	0.001	0.022	0.022	0.940	0.002	0.022	0.022	0.940
For the time to IBTR								
$\beta_1 = 0.5$	0.003	0.124	0.117	0.946	-0.112	0.099	0.093	0.752
$\beta_2 = 0.5$	-0.004	0.251	0.235	0.936	-0.105	0.211	0.203	0.903
$\zeta = 3$	0.053	0.378	0.355	0.936	-0.316	0.262	0.262	0.769
For the time to terminal event								
$\eta_1 = 0.5$	0.026	0.139	0.149	0.958	-0.207	0.066	0.065	0.122
$\eta_2 = 0.5$	0.012	0.261	0.264	0.953	-0.215	0.142	0.135	0.645
$\gamma = 0.5$	0.026	0.262	0.262	0.880				

SE, standard error; SEM, mean of the standard error estimates; CP, coverage probability; IBTR, ipsilateral breast tumor relapse; NP, new ipsilateral primary tumor.

with slightly larger bias, SE, and SEM. Under model overparameterization, the estimate of γ from the joint model is correctly close to zero, although the SE estimate is slightly inflated.

Table II suggests that in setting II with some correlation (i.e., $\gamma = 0.5$), the joint model provides less biased estimates of covariate effects on time to IBTR (β) and time to terminal event (η) than the reduced model. Although the SEs of the parameter estimates are similar for the parameters predicting the probability of IBTR (α), when events are assumed to be independent, the SEs of covariate effects related to the two event processes (i.e., β and η) are drastically underestimated, as would be expected under the incorrect assumption of independence of event times. Correspondingly, the estimated coverage probabilities are nearer to the 0.95 nominal level for the joint model versus the reduced model. For parameters related to the times to IBTR and terminal event (i.e., β and η), the coverage probabilities produced by the reduced model are extremely inaccurate. Further, the parameter estimate of ζ , which represents the additional log hazard of being TR versus NP, is highly negatively biased when fitting the reduced model, because the dependency of the events cannot be accurately measured by assuming independent likelihoods. In comparison, the true value of ζ in simulation is adequately recovered by the joint model.

The proposed joint modeling framework is a shared random effects model, in which two survival models share a function of the random effects. This model has been widely used in joint model literature. Another alternative is multivariate random effects survival models, which assume that two survival processes are linked by a vector of two correlated random effects. Specifically, we expressed models (2) and (3) as $\lambda(t_D) = \lambda_0(t_D) \exp(\mathbf{x}_D \boldsymbol{\eta} + u_1)$ and $r^N(t) = r_0(t) \exp(\mathbf{x}_R \boldsymbol{\beta} + u_2)$, respectively, where $\mathbf{u} = (u_1, u_2)' \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ being denoted as $((\sigma_1^2, \rho\sigma_1\sigma_2), (\rho\sigma_1\sigma_2, \sigma_2^2))$. However, this model is unidentifiable in this context because each breast cancer patient can possibly have only one IBTR occurrence and one breast cancer death. There is insufficient information to identify parameters in the multivariate random effects correlation matrix. Nevertheless, it is essential to assess the robustness of the proposed joint model under random effects structure misspecification. We simulate 1000 data sets from the multivariate random effects survival model with $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, and $\rho = 0.636$ and all other parameters identical to setting II. Table III compares the simulation results from the joint and reduced models. The results indicate that the proposed joint model provides reasonable results to all parameters with relatively small bias, SEM being close to SE, and CP being around the nominal value. In contrast, the reduced model gives biased estimates and low CP, especially for α , β , η , and ζ , although the estimation results for the sensitivity p and specificity q are reasonably good.

The simulation study suggests that the joint model provides results comparable with the reduced model under the independent terminal event (setting I). Under the dependent terminal mechanism (setting II), the joint model provides more accurate estimates for the regression parameters of the

Table III. Simulation results of fitting the proposed joint model and the reduced model where data are simulated from the multivariate random effects survival model.

Parameter	Joint model				Reduced model			
	Bias	SE	SEM	CP	Bias	SE	SEM	CP
For the probability of IBTR being NP								
$\alpha_0 = -2$	-0.062	0.436	0.389	0.906	0.117	0.411	0.380	0.872
$\alpha_1 = 3$	0.092	0.560	0.514	0.918	-0.115	0.547	0.503	0.875
$\alpha_2 = 4$	0.140	0.794	0.728	0.928	-0.162	0.744	0.719	0.896
$p = 0.8$	0.008	0.069	0.066	0.908	-0.004	0.070	0.069	0.931
$q = 0.9$	0.001	0.022	0.022	0.942	0.000	0.023	0.022	0.929
For the time to IBTR								
$\beta_1 = 0.5$	-0.002	0.135	0.122	0.930	-0.122	0.093	0.092	0.721
$\beta_2 = 0.5$	0.012	0.260	0.241	0.935	-0.118	0.21	0.196	0.883
$\zeta = 3$	-0.005	0.428	0.400	0.917	-0.526	0.26	0.253	0.442
For the time to terminal event								
$\eta_1 = 0.5$	0.010	0.111	0.114	0.954	-0.111	0.071	0.066	0.608
$\eta_2 = 0.5$	0.015	0.195	0.197	0.960	-0.112	0.134	0.136	0.875

SE, standard error; SEM, mean of the standard error estimates; CP, coverage probability; IBTR, ipsilateral breast tumor relapse; NP, new ipsilateral primary tumor.

probability of IBTR and two survival times than the reduced model. Moreover, when the random effects structure is misspecified, the proposed joint model still provides accurate estimates for all parameters of interest, while the reduced model gives biased estimates to most parameters.

5. Real data analysis

In this section, we applied the methods to the breast cancer patient data set ($n = 4477$). We defined the terminal event as the breast cancer-related death (251 patients); all other terminal events (e.g., death unrelated to breast cancer, censoring) are treated as independent censoring. The following three covariates are considered: x_1 represents age at breast cancer diagnosis (mean: 55.4 years; SD: 12.1 years); x_2 represents whether a distant recurrence developed in organs other than breasts (e.g., bones, lung, brain, and liver; prevalence: 7.5%); x_3 represents primary tumor stage (1 if more aggressive stage II or higher and 0 otherwise; prevalence: 23.1%) [39]. Let $\mathbf{x}_P = (1, x_1, x_2, x_3)$ and $\mathbf{x}_R = \mathbf{x}_D = (x_1, x_3)$.

We use the adaptive Gaussian quadrature estimation method. To specify the baseline hazard functions $\lambda_0(t_D)$ and $r_0(t)$, we divide the time to IBTR and time to breast cancer death into M_1 and M_2 intervals by every $1/M_1$ th and $1/M_2$ th quantiles, respectively. We adopt $M_1 = M_2 = 4$ in this article. We have also explored other selections of M_1 and M_2 , and the results are very similar. For model selection and comparison, the Akaike information criterion (AIC) [40] and Bayesian information criterion (BIC) are computed [41]. The joint model performs significantly better than the reduced model with smaller AIC (7578.0 vs 7632.6) and BIC (7687.0 vs 7728.7).

Table IV provides the means, SE, p -values, and 95% CI of the parameters from the proposed joint model (6). Covariate interpretation is as follows. The odds ratio of the IBTR being NP for the patients who developed distant recurrence is 0.012 (i.e., $\exp(-4.435)$; 95% CI: $[4.273e - 4, 0.329]$) versus the patients who did not develop distant recurrence. The covariate age at diagnosis and tumor stage are not statistically significant. These findings are consistent with the results of previous studies [28–30]. The sensitivity and specificity estimates of the diagnostic test are 0.648 (95% CI: $[0.494, 0.802]$) and 0.873 (95% CI: $[0.757, 0.988]$), respectively. Thus, this diagnostic test is more likely to correctly classify TR patients than NP patients. Because TR patients tend to have shorter time from BCT to IBTR and need more aggressive treatment than NP patients, the misclassification of TR patients into NP is likely to be more costly than vice versa.

We interpreted the covariate effects on time to IBTR as follows. The hazard rate of IBTR is 0.588 (i.e., $\exp(-0.53)$; 95% CI: $[0.522, 0.664]$) for every 10-year increase in the covariate age at breast cancer diagnosis. The TR status significantly increases the hazard of IBTR, with an hazard rate of 6.209 (i.e., $\exp(1.826)$; 95% CI: $[3.127, 12.317]$) compared with NP patients. This large difference in hazard between TR and NP statuses is visually identifiable by the Kaplan–Meier curves (Figure 2) and

Table IV. The maximum likelihood estimates, standard errors, p -values, and 95% confidence interval of the parameters from the proposed joint model.

	MLE	SE	p	95% CI
For the probability of IBRT being NP				
Intercept	4.298	2.511	0.087	[-0.626, 9.222]
Age	-0.006	0.031	0.855	[-0.066, 0.055]
Distant	-4.435	1.695	0.009	[-7.758, -1.111]
Stage	-0.730	0.898	0.417	[-2.490, 1.031]
p	0.648	0.079		[0.494, 0.802]
q	0.873	0.059		[0.757, 0.988]
For the time to IBTR				
Age	-0.053	0.006	< 0.001	[-0.065, -0.041]
Stage	-0.015	0.171	0.928	[-0.350, 0.319]
ζ	1.826	0.350	< 0.001	[1.140, 2.511]
For the time to breast cancer death				
Age	-0.027	0.006	< 0.001	[-0.039, -0.015]
Stage	1.088	0.149	< 0.001	[0.796, 1.380]
γ	0.532	0.144	< 0.001	[0.249, 0.815]
σ_u^2	1.592	0.433		

MLE, maximum likelihood estimates; SE, standard error; CI, confidence interval; IBTR, ipsilateral breast tumor relapse; NP, new ipsilateral primary tumor.

has also been reported previously [28, 29, 42]. Finally, the hazard rate of breast cancer death is 0.763 (i.e., $\exp(-0.27)$; 95% CI: [0.677, 0.861]) for every 10-year increase in age at breast cancer diagnosis. The hazard rate of breast cancer death is 2.968 (i.e., $\exp(1.088)$; 95% CI: [2.217, 3.975]) for the patients with higher tumor stage as compared with the ones with lower tumor stage. For model association, note that $\hat{\gamma}$ is significantly larger than zero (mean: 0.532, $p < 0.001$, 95% CI: [0.249, 0.815]), indicating that patients with larger hazard of breast cancer death tend to have IBTR earlier. The statistical significance of $\hat{\gamma}$ also suggests that the time to death is strongly correlated to the time to relapse, which justifies the need of joint modeling.

6. Discussion

This article presents a joint analysis method to model the binomial regression with a single misclassified binary outcome and the correlated time to IBTR, subject to a dependent terminal event and in the absence of a gold standard. The models of time to IBTR and time to the terminal event are linked together via shared random effects. The simulation study shows that when the terminal event is independent of the relapse process, the proposed joint model provides comparable results to the reduced model. However, when there exists a dependent terminal event, the joint model provides accurate estimates of all parameters, while the reduced model gives severely biased estimates for the regression parameters for the probability of NP at IBTR, and the two survival times. The analysis of the breast cancer data set indicates that the target diagnostic test has higher specificity than sensitivity (i.e., more likely to correctly classify TR patients than NP patients). This is preferable because TR patients tend to have shorter time to IBTR and need more aggressive treatment than NP patients, and the misclassification of TR patients into NP is likely to be more costly than vice versa. Further, our analysis results provide useful information in identifying and quantifying the covariate effects on the probability of IBTR status and survival times. From the analysis results, we found that the sensitivity and specificity of the diagnostic test, which uses only clinical and pathological criteria, are low. A more accurate diagnostic test using the molecular criteria should be developed in the future, as pointed out by Huang *et al* [28]. We can broadly apply the proposed method to many studies with similar data structure consisting of misclassified binary outcome and a dependent terminal event. We can conveniently fit the proposed joint model using adaptive Gaussian quadrature tools implemented in SAS procedure NLMIXED and can be easily accessible to, modified, and extended by applied researchers.

The difficulty of estimating diagnostic accuracy without a gold standard has been discussed in the literature. For example, Albert and Dodd [43] showed that the estimation of diagnostic accuracy and

prevalence is sensitive to the choice of dependence structure for studies with multiple diagnostic tests. They hence recommended collecting gold standard information on a fraction of subjects if possible, doing a sensitivity analysis using very different methods for accounting for multiple test dependence, and taking a large number of tests. However, there is only a single imperfect diagnostic test in this article. It requires information on the misclassification structure to make the model identifiable [22]. The covariates in x_p in modeling IBTR status in model (1) are instrumental variables, and they make the sensitivity and specificity parameters identifiable when their number of different possible realizations is sufficient [23]. Moreover, 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. The detailed discussion of the identifiability, modeling, and parameter estimation of multiple conditional dependent diagnostic tests can be found in some recent literature, for example, Dendukuri and Joseph [44], Georgiadis *et al* [45], and Jones *et al* [46].

The joint modeling strategy has several limitations that can be addressed in future research. One limitation is the non-differential assumption; that is, the sensitivity and specificity do not depend on the covariates. This assumption has been relaxed in some statistical literature [47]. We will explore this extension in the future research. In addition, this article only considers a single type terminal event. In the presence of multiple failure types, for example, death due to breast cancer and other causes, the proposed joint model can be extended to accommodate competing risks survival data. One alternative to the proposed joint model is the models based on pattern mixture modeling that stratifies the subjects into subgroups by their censoring patterns, and then models the stratified subgroups separately, for example, Hogan and Laird [48, 49], Molenberghs and Verbeke [50], Zhang *et al* [51]. We would like to investigate the application of the pattern mixture modeling to the proposed joint modeling framework. Another issue is the normality assumption of random effects in our joint model. Some researchers [52, 53] have reported that the statistical inference of joint models is generally robust to the departure from the normality assumption. It is of interest to investigate our joint model's performance when the underlying random effects distribution is symmetric non-normal or even asymmetric. Moreover, we assumed the random effects variance to be homogeneous (same for all individuals). However, the random effects variance may depend on subject-specific characteristics and is thus heterogeneous. Ignoring the heterogeneity can result in biased estimates [54, 55]. As a future direction, we would address the non-normal and heterogeneous random effects issues in the proposed joint modeling framework.

Appendix A

```

/* Invoke proc nlmixed using adaptive Gaussian Quadrature */
proc nlmixed data=piecewise hess maxit=8000 MAXFUNC=8000;
  /* define the array */
  * piecewise constant function for death
  array r_D{3} r_D1-r_D3;
  * cut points for the intervals of death event
  array tau_D{4} tau_D1-tau_D4;
  * piecewise constant function for IBTR
  array r_R{3} r_R1-r_R3;
  * cut points for the intervals of IBTR
  array tau_R{4} tau_R1-tau_R4;
  * initialize parameters
  parms se=0.7, sp=0.7, sigma2_u=1.5, zeta=2, gamma=0.3,
        r_D1=0.5, r_D2=0.5, r_D3=0.3, r_R1=0.8, r_R2=0.8, r_R3=0.6;
  * set boundaries for parameters
  bounds l>se>0, l>sp>0, sigma2_u>0, r_D1 r_D2 r_D3 r_R1 r_R2 r_R3 >0;
  * specify the cut points for the intervals
  tau_R[1]=0; tau_R[2]=2.5; tau_R[3]=8; tau_R[4]=20;
  tau_D[1]=0; tau_D[2]=2.5; tau_D[3]=8; tau_D[4]=20;
  *Calculate the cumulative hazard for relapse and death processes;
  G_D=0; G_R=0; G_DR=0;
  do i=1 to 3;
    if i<IndicatorD then G_D=G_D+r_D[i]*(tau_D[i+1]-tau_D[i]);
    if i<IndicatorR then G_R=G_R+r_R[i]*(tau_R[i+1]-tau_R[i]);
    if i<IndicatorDR then G_DR=G_DR+r_R[i]*(tau_R[i+1]-tau_R[i]);
    if i=IndicatorD then G_D=G_D+(T_D-tau_D[i])*r_D[i];
    if i=IndicatorR then G_R=G_R+(T_R-tau_R[i])*r_R[i];
    if i=IndicatorDR then G_DR=G_DR+(T_D-tau_R[i])*r_R[i];
  end;

```

```

eta_p=alpha0+alpha1*x1;
eta_D=etal*x1+u;
eta_TR=beta1*x1+gamma+u+zeta;
eta_NP=beta1*x1+gamma*u;
pi=exp(eta_p)/(1+exp(eta_p));
lambda_tD=r_D[IndicatorD]*exp(eta_D);
r_TR=r_R[IndicatorR]*exp(eta_TR);
r_NP=r_R[IndicatorR]*exp(eta_NP);
S_D=exp(-exp(eta_D)*G_D);
S_TR= exp(-exp(eta_TR)*G_R);
S_NP= exp(-exp(eta_NP)*G_R);
S_TRD= exp(-exp(eta_TR)*G_DR);
S_NPD= exp(-exp(eta_NP)*G_DR);
*Calculate the log likelihood;
if R eq 1 then do;
  if z eq 1 then ll=log(pi*se*r_NP*S_NP+(1-pi)*(1-sp)*r_TR*S_TR)+delta*log(lambda_tD)+log(S_D);
  if z eq 0 then ll=log(pi*(1-se)*r_NP*S_NP+(1-pi)*sp*r_TR*S_TR)+delta*log(lambda_tD)+log(S_D);
end;
if R eq 0 then ll=log(pi*S_NPD+(1-pi)*S_TRD)+delta*log(lambda_tD)+log(S_D);
*specify random effect;
random u~normal(0,sigma2_u) subject=id;
model z~general(ll);
run;

```

Acknowledgements

The National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant KL2 TR000370, supported the project described. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The authors thank the editor and an anonymous associate editor for their insightful comments and suggestions that led to a marked improvement of the article.

References

1. Carroll RJ, Ruppert D, Stefanski L, Crainiceanu C. *Measurement Error in Nonlinear Models: A Modern Perspective*. Vol. 105. Chapman and Hall/CRC: Boca Raton, FL, 2010.
2. Chen B, Grace Y, Cook R. Weighted generalized estimating functions for longitudinal response and covariate data that are missing at random. *Journal of the American Statistical Association* 2010; **105**(489):336–353.
3. Neuhaus J. Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika* 1999; **86**:843–855.
4. Neuhaus J. Analysis of clustered and longitudinal binary data subject to response misclassification. *Biometrics* 2002; **58**(3):675–683.
5. Barron B. The effects of misclassification on the estimation of relative risk. *Biometrics* 1977; **33**:414–418.
6. Greenland S. The effect of misclassification in matched-pair case-control studies. *American Journal of Epidemiology* 1982; **116**:402–406.
7. Greenland S, Ericson C, Kleinbaum D. Correcting for misclassification in two-way tables and matched-pair studies. *International Journal of Epidemiology* 1983; **12**:93–97.
8. Magder L, Hughes J. Logistic regression when the outcome is measured with uncertainty. *American Journal of Epidemiology* 1997; **146**:195–203.
9. Spiegelman D, Rosner B, Logan R. Estimation and inference for logistic regression with covariate misclassification and measurement error in main study/validation study designs. *Journal of the American Statistical Association* 2000; **95**:51–61.
10. Robins J, Rotnitzky A, Zhao L. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 1994; **89**:846–866.
11. Prescott G, Garthwaite P. A Bayesian approach to prospective binary outcome studies with misclassification in a binary risk factor. *Statistics in Medicine* 2005; **24**:3463–3477.
12. Carroll R, Ruppert D, Stefanski L, Crainiceanu C. *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and Hall/CRC: Boca Raton, FL, 2006.
13. Sheps S, Schechter M. The assessment of diagnostic tests. *Journal of the American Medical Association* 1984; **17**:2418–2422.
14. Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. *American Journal of Epidemiology* 1993; **137**:1251–1258.
15. Albert P, Dodd L. On estimating diagnostic accuracy from studies with multiple raters and partial gold standard evaluation. *Journal of the American Statistical Association* 2008; **103**:61–73.
16. Hui S, Zhou X. Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research* 1998; **7**:354–370.
17. Luo S, Yi M, Huang X, Hunt K. A Bayesian model for misclassified binary outcomes and correlated survival data with applications to breast cancer. *Statistics in Medicine* 2013; **32**(13):2320–34.
18. Rindskopf D, Rindskopf W. The value of latent class analysis in medical diagnosis. *Statistics in Medicine* 1986; **5**:21–27.

19. Formann A. Measurement errors in caries diagnosis: some further latent class models. *Biometrics* 1994; **50**:865–871.
20. Duffy S, Warwick J, Williams A, Keshavarz H, Kaffashian F, Rohan T, Nili F, Sadeghi-Hassanabadi A. A simple model for potential use with a misclassified binary outcome in epidemiology. *Journal of epidemiology and community health* 2004; **58**(8):712–717.
21. Joseph L, Gyorkos T, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology* 1995; **141**(3):263–272.
22. Ren D, Stone R. A Bayesian adjustment for covariate misclassification with correlated binary outcome data. *Journal of Applied Statistics* 2007; **34**:1019–1034.
23. Nagelkerke N, 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**:739–44.
24. Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* 1996; **15**(15):1663–1685.
25. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**(4):465–480.
26. Rouzier R, Extra J, Carton M, Falcou M, Vincent-Salomon A, Fourquet A, Pouillart P, E B. 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**:3828–35.
27. Chen A, Meric-Bernstam F, Hunt K, Thames H, Oswald M, Outlaw E, Strom E, McNeese M, Kuerer H, Ross M, Singletary S, Ames F, Feig B, Sahin A, Perkins G, Schechter N, Hortobagyi G, Buchholz T. Breast conservation after neoadjuvant chemotherapy: the MD Anderson Cancer Center experience. *Journal of Clinical Oncology* 2004; **22**:2303–12.
28. Huang E, Buchholz T, Meric F, Krishnamurthy S, Mirza N, Ames F, Feig B, Kuerer H, Ross M, Singletary S, McNeese M, Strom E, Hunt K. 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**:2059–67.
29. 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**:104–11.
30. Abd-Alla H, Lotayef M, Abou Bakr A, Moneer M. Ipsilateral in-breast tumor relapse after breast conservation therapy: true recurrence versus new primary tumor. *Journal of the Egyptian National Cancer Institute* 2006; **18**:183–90.
31. Liu L, Conaway M, Knaus W, Bergin J. A random effects four-part model, with application to correlated medical costs. *Computational Statistics & Data Analysis* 2008; **52**(9):4458–4473.
32. 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.
33. Liu L, Strawderman R, Cowen M, Shih Y. A flexible two-part random effects model for correlated medical costs. *Journal of Health Economics* 2010; **29**(1):110–123.
34. Lesaffre E, Spiessens B. On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2001; **50**(3):325–335.
35. Lawless J, Zhan M. Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics* 1998; **26**(4):549–565.
36. Feng S, Wolfe R, Port F. 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.
37. 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.
38. He B, Luo S. Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson's disease. *Statistical Methods in Medical Research* 2013. published online on April 16, 2013, PMID: PMC3883896, DOI: 10.1177/0962280213480877.
39. Singletary S, Allred C, Ashley P, Bassett L, Berry D, Bland K, Borgen P, Clark G, Edge S, Hayes D, Hughes L, Hutter R, Morrow M, Page D, Recht A, Theriault R, Thor A, Weaver D, Wieand H, Greene F. Revision of the American Joint Committee on Cancer staging system for breast cancer. *Journal of Clinical Oncology* 2002; **20**(17):3576–7.
40. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974; **19**(6):716–723.
41. Schwarz G. Estimating the dimension of a model. *The Annals of Statistics* 1978; **6**(2):461–464.
42. Smith T, Lee D, Turner B, Carter D, Haffty B. 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.
43. Albert PS, Dodd LE. A cautionary note on the robustness of latent class models for estimating diagnostic error without a gold standard. *Biometrics* 2004; **60**(2):427–435.
44. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* 2001; **57**(1):158–167.
45. Georgiadis M, Johnson W, Gardner I, 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.
46. Jones G, Johnson W, Hanson T, Christensen R. Identifiability of models for multiple diagnostic testing in the absence of a gold standard. *Biometrics* 2010; **66**(3):855–863.
47. Paulino C, Soares P, Neuhaus J. Binomial regression with misclassification. *Biometrics* 2003; **59**(3):670–675.
48. Hogan JW, Laird NM. Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* 1997; **16**:239–257.

49. Hogan JW, Laird NM. Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* 1997; **16**:259–272.
50. Molenberghs G, Verbeke G. *Models for Discrete Longitudinal Data*. Springer Verlag: New York, 2005.
51. Zhang S, Müller P, Do KA. A Bayesian semiparametric survival model with longitudinal markers. *Biometrics* 2010; **66**(2):435–443.
52. Song X, Davidian M, Tsiatis AA. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* 2002; **58**(4):742–753.
53. Zeng D, Cai J. Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *The Annals of Statistics* 2005; **33**(5):2132–2163.
54. Heagerty PJ, Kurland BF. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* 2001; **88**(4):973.
55. Daniels MJ, Zhao YD. Modelling the random effects covariance matrix in longitudinal data. *Statistics in Medicine* 2003; **22**(10):1631–1647.