Published in final edited form as:

J Appl Stat. 2015 ; 42(5): 1080–1090. doi:10.1080/02664763.2014.995606.

Simultaneous inference of a misclassified outcome and competing risks failure time data

Sheng Luo [Assistant Professor],

Division of Biostatistics, The University of Texas Health Science Center at Houston, 1200 Pressler St, Houston, TX 77030, USA (sheng.t.luo@uth.tmc.edu; Phone: 713-500-9554)

Xiao Su, PhD [candidate],

Division of Biostatistics, The University of Texas Health Science Center at Houston

Min Yi [database manager], and

Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Texas 77030, USA

Kelly K. Hunt [Professor]

Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Texas 77030, USA

Abstract

Ipsilateral breast tumor relapse (IBTR) often occurs in breast cancer patients after their breast conservation therapy. The IBTR status' classification (true local recurrence versus new ipsilateral primary tumor) is subject to error and there is no widely-accepted gold standard. Time to IBTR is likely informative for IBTR classification because new primary tumor tends to have a longer mean time to IBTR and is associated with improved survival as compared with the true local recurrence tumor. Moreover, some patients may die from breast cancer or other causes in a competing risk scenario during the follow-up period. Because the time to death can be correlated to the unobserved true IBTR status and time to IBTR (if relapse occurs), this terminal mechanism is non-ignorable. In this article, we propose a unified framework that addresses these issues simultaneously by modeling the misclassified binary outcome without a gold standard and the correlated time to IBTR, subject dependent competing terminal events. We evaluate the proposed framework by a simulation study and apply it to a real dataset consisting of 4,477 breast cancer patients. The adaptive Gaussian quadrature tools in SAS procedure NLMIXED can be conveniently used to fit the proposed model. We expect to see broad applications of our model in other studies with a similar data structure.

Keywords

Diagnostic error; Imperfect tests; Binary outcome; Latent class model; Sensitivity; Specificity

1 Introduction

Breast conservation therapy (BCT) is a breast conserving surgery followed by moderate-dose radiation therapy to eradicate the microscopic residual disease and it offers similar overall survival and disease-free survival rates as does mastectomy in breast cancer patients¹. However, approximately 8–20% of patients undergoing BCT for primary breast cancer will experience ipsilateral breast tumor relapse (IBTR), defined as tumor relapse in the treated breast². The IBTR status is classified as either true local recurrence tumor (TR, residual cancer cells grow gradually to detectable size) or new ipsilateral primary tumor (NP, first occurrence of cancer independently arising in the preserved breast) by some diagnostic tests. Because of the uncertainties of clinical and pathologic criteria for classification, the diagnostic tests for IBTR classification are error-prone and there is no widely-accepted gold standard. The correct classification of IBTR status is of clinical significance in treatment planning and therapeutic decision-making. For example, more aggressive chemotherapy and hormone therapy may be necessary for TR patients, while milder treatment may be sufficient for NP patients.

Diagnostic accuracy of medical tests or raters has attracted considerable research interest in many public health and biomedical fields³. In the presence of a gold standard on all individuals, the estimation of diagnostic accuracy in terms of sensitivity and specificity is straightforward. Unfortunately, gold standard evaluation may be expensive, time-consuming, or unethical to perform on all subjects⁴. If the gold standard is only available to a subset of subjects, the techniques based on likelihood or weighted estimating equations have been proposed^{5,6}. When a gold standard does not exist and multiple diagnostic tests are available, it has been shown that if the conditional independence assumption (given true disease status, the statuses of all tests are independent) is incorrectly assumed, parameter estimates may be biased⁷. When the disease status is binary, Albert and Dodd⁸ showed that the choice of dependence structure for multiple diagnostic tests influences the estimation of diagnostic accuracy and prevalence. In the presence of an imperfect reference standard (the best available reference tests are still subject to error, but have high sensitivity and specificity), Albert⁹ showed that the inference on diagnostic accuracy are robust to misspecification of conditional dependence between tests. This research has been extended by using the additional information on a set of four international experts¹⁰. However, in the scenario of a single diagnostic test and no additional information (e.g., imperfect reference standard, partial gold standard), all aforementioned models are not directly applicable.

In the presence of a single imperfect diagnostic test, some additional information on the misclassification structure is required in order to make the model identifiable and to adjust for potential bias due to misclassification¹¹. To this end, Nagelkerke et al.¹² suggested modeling the unobserved true disease status as a function of an instrumental variable, which is an additional parameter to increase the outcome degrees of freedom. This instrumental variable framework can be extended to utilize the available correlated survival time to gain additional information regarding the disease status classification. For example, in breast cancer, time to IBTR is potentially informative for IBTR status classification because NP patients has a longer mean time to IBTR and is associated with improved survival^{13,14}. Hence we would like to incorporate this correlated survival information when modeling the

misclassified IBTR status. Moreover, there are several additional complicated issues. First, some patients died from breast cancer during the follow-up period, in addition to the independent censoring such as moving to other cities. Because the disease-related death can be correlated to the unobserved true IBTR status and time to IBTR (if relapse occurs), this terminal mechanism is non-ignorable¹⁵. Second, patients may die from other causes, producing a competing risk scenario. It is critically important to include the effect of competing risks in the modeling framework because these competing risks are highly correlated with IBTR occurrence. The literature in joint modeling suggests that failure to account for the non-ignorable terminal mechanism leads to biased estimates¹⁶. However, most joint models developed in the literature have dealt with univariate time-to-event data and proposed a submodel for survival data with a single failure type, which is not applicable in the scenario of competing risks. To study the covariate effects in the competing risk settings, Cox analysis of cause-specific hazards has been widely used. The joint modeling of longitudinal and competing risk data employs the cause-specific hazard sub-model with a latent association between longitudinal measurements and each cause of failure¹⁷.

In this article, we provide a unified framework that accounts for all the issues together. We proposed the model framework for a binary outcome classified by a single imperfect diagnostic test and the correlated time to IBTR, in the presence of two competing risks, i.e., breast cancer death and other cause death. The article proceeds as follows. Section 2 describes the dataset that motivates this research and presents the statistical methodology. Section 3 presents the simulation results. Section 4 presents results of the motivating dataset analysis. Section 5 gives a discussion.

2 Data and Statistical models and likelihood inference

2.1 Motivating Data Set

The motivation of this statistical research comes from a clinical dataset consisting of 4, 477 patients with invasive breast cancer who underwent BCT at The University of Texas MD Anderson Cancer Center (MDACC). The collected variables include patient characteristics (age, ethnicity, family history of breast and other cancers), treatment characteristics (surgery and radiation), and tumor characteristics (location, histology, stage, and size). In this dataset, there are 397 breast cancer patients who developed IBTR. By using a single imperfect diagnostic test based on tumor location and histologic subtype (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; NP otherwise¹⁸), 201 (50.6%) and 196 (49.4%) of the 397 patients with IBTR were classified as TR and NP, respectively. Among all patients, 251 (5.6%) died from breast cancer (referred to as breast cancer death) and 212 (4.7%) died from other causes (referred to as other death).

In this motivating breast cancer dataset, time to IBTR is likely informative for IBTR status classification because NP relapse has a longer mean time to IBTR and is associated with improved survival as compared with TR tumor relapse¹³. To visualize the correlation between time to IBTR and the possibly misclassified IBTR status, Figure 1 displays the curves of cumulative incidence of IBTR (sample size $N = 397$). The plot suggests that TR

patients have larger probability of having IBTR as compared with NP patients ($p < 0.0001$ in tests by Gray¹⁹).

Furthermore, the occurrence and status of IBTR are likely informative for the death from breast cancer and other causes. To visualize this correlation, Figure 2 displays the curves of cumulative incidences of breast death (left panel) and other death (right panel) for patients with IBTR classified as TR or NP, and for patients without IBTR. The left panel suggests that the TR patients have much larger probability of dying from breast cancer than all other patients, while there is no significance difference among the patients with NP and those without IBTR. The right panel suggests that the patients without IBTR have much larger probability of dying from other causes than the patients with IBTR, while there is no significance difference among the TR and NP patients. The tests by Gray¹⁹ suggest significance difference with $p < 0.0001$ in both panels. Hence, the occurrence and classification of IBTR is strongly correlated with the breast cancer death and other death. The purpose of this article is to develop a unified modeling framework of a misclassified binary outcome and time to relapse with competing dependent terminal event in order to (1) quantify covariate effects on the probability of IBTR being NP, on the hazards of IBTR, and on the hazards of the competing terminal events, (2) estimate the sensitivity and specificity of the diagnostic test, (3) study the association among all survival times.

2.2 Likelihood formulation

Let variable R be the indicator for 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 death time D (e.g., due to breast cancer or other causes), both starting from the time of BCT. Each patient may experience one of the K ($K = 2$ in the motivating example) distinct failure types or could be right censored. Let event indicator d take a value in $\{0, 1, \dots, K\}$, with $d = 0$ representing an independent censored event and $d = k$ representing failure from the k th type of failure, where $k = 1, \dots, K$. Let $t = \min(t_R, t_D)$. Thus, we have $t = t_R$ and $R = 1$ for the patients with IBTR and $t = t_D$ and $R = 0$ for the patients without IBTR. To illustrate this complex data structure, Figure 3 displays all possible events. In terms of IBTR occurrence, there are 397 patients with IBTR and 4,080 patients without IBTR. In term of failure types, there were 4,014 patients who were censored ($d = 0$), 251 breast cancer death ($d = 1$), and 212 other death ($d = 2$).

Let variable y (1 if NP, 0 if TR) be the unobserved true IBTR status. Let z (1 if NP, 0 if TR) be the observed outcome of y from the imperfect diagnostic test. Let p and q be the sensitivity and specificity of the test given y , respectively, i.e., $p = p(z = 1|y = 1)$ and $q = p(z = 0|y = 0)$. We make the following assumptions: (a) non-differential assumption (sensitivity and specificity do not depend on covariates); (b) conditional on the shared random effects u and covariates, the diagnostic test result z , time to IBTR t_R , and time to breast cancer or other death t_D are independent. The covariate vectors $\mathbf{x}_p$, $\mathbf{x}_R$, and $\mathbf{x}_D$ are for the probability of IBTR being NP, time to IBTR, and competing risks of death, respectively, and they can share part of or all the covariates. To model the probability of the IBTR being NP, the binomial regression model is

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

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

The hazard function of time to IBTR is

$$r^N(t) = r_0(t) \exp(\mathbf{x}_R \boldsymbol{\beta} + \gamma u), \quad (2)$$

where the NP status is denoted by the superscript N , the baseline hazard function for IBTR is $r_0(t)$, and the random effects $u \sim N(0, \sigma_u^2)$. The corresponding survival function $S^N(t) = \exp[-\exp(\mathbf{x}_R \boldsymbol{\beta} + \gamma u) \int_0^t r_0(x) dx]$. The hazard for TR patients is $r^T(t) = r_0(t) \exp(\mathbf{x}_R \boldsymbol{\beta} + \zeta + \gamma u)$, where the additional hazard of being TR patients compared with NP patients is denoted by ζ and the TR status is denoted by the superscript T . The likelihood of observing IBTR at time $t = \min(t_R, t_D)$ is

$$L_1(\boldsymbol{\theta}_1|u) = \left\{ [\pi p r^N(t) S^N(t) + \pi q r^T(t) S^T(t)]^z [\pi p r^N(t) S^N(t) + \pi q r^T(t) S^T(t)]^{\bar{z}} \right\}^R \cdot \left\{ [\pi S^N(t) + \pi S^T(t)] \right\}^{1-R}, \quad (3)$$

where the parameter vector $\boldsymbol{\theta}'_1 = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \gamma, \sigma_u, \zeta, p, q)$.

A cause-specific hazard model with random effects is assumed for the competing risks survival data:

$$\lambda_k(t_D) = \lambda_{0k}(t_D) \exp(\mathbf{x}_D \boldsymbol{\eta}_k + \nu_k u), \quad k=1, \dots, K, \quad (4)$$

where the function $\lambda_k(t_D)$ is the instantaneous hazard rate from cause k at time t_D , given the regression vector $\mathbf{x}_D$ and the random effects u , in the presence of other competing terminal events. The parameter vector $\boldsymbol{\eta}_k$ represents the effects of the covariate vector $\mathbf{x}_D$ to the k th type of terminal event. The parameter ν_k represents the effects of the random effects u with $\nu_1 = 1$ to ensure identifiability. The corresponding survival function for the k th competing

terminal event is $S_k(t_D) = \exp[-\int_0^{t_D} \lambda_{0k}(t) \exp(\mathbf{x}_D \boldsymbol{\eta}_k + \nu_k u) dt]$. The likelihood of the competing terminal events is

$$L_2(\boldsymbol{\theta}_2|u) = \prod_{k=1}^K \left\{ [\lambda_{0k}(t_D)]^{I(d=k)} S_k(t_D) \right\}, \quad (5)$$

where the parameter vector $\boldsymbol{\theta}'_2 = (\boldsymbol{\eta}', \nu_2, \dots, \nu_K)$ with $\boldsymbol{\eta}' = (\boldsymbol{\eta}'_1, \dots, \boldsymbol{\eta}'_K)$.

The marginal likelihood for one patient is

$$L(\boldsymbol{\theta}) = \int L(\boldsymbol{\theta}|u) f(u) du = \int L_1(\boldsymbol{\theta}_1|u) L_2(\boldsymbol{\theta}_2|u) f(u) du, \quad (6)$$

where $f(u)$ is normal distribution $N(0, \sigma_u^2)$ and the parameter vector $\theta' = (\theta'_1, \theta'_2)$. There are two sources of correlation in the proposed model. The strength of association between IBTR and the competing risks is measured by the parameter γ and its significance can be tested by the hypothesis $H_0 : \gamma = 0$. The strength of association between different failure types (e.g., breast cancer and other death) is measured by the parameter v_k and its significance can be tested by the hypothesis $H_0 : v_k = 0$.

2.3 Maximum likelihood estimation

The marginal likelihood in model (6) involves unspecified baseline hazard functions $r_0(t)$ and $\lambda_{0k}(t_D)$. We adopt a piecewise constant function to approximate the baseline hazard functions. It has been shown that survival models using a piecewise constant baseline hazard function with fixed cut points yield good estimators for both fixed and random effects²⁰. Given a set of fixed time points $\tau_R = (\tau_{R0}, \dots, \tau_{Rm})$ with the constraint $0 = \tau_{R0} < \tau_{R1} < \dots < \tau_{Rm}$, and the baseline hazard vector $\mathbf{g}_R = (g_{R0}, g_{R1}, \dots, g_{Rm-1})$ for the time to IBTR t_R , the piecewise constant baseline hazard function is defined as $r_0(t) = \sum_{l=0}^{m-1} g_{Rl} I_l(t)$, with indicator function $I_l(t) = 1$ if $\tau_l \leq t < \tau_{l+1}$ and 0 otherwise. Similarly, the piecewise constant baseline hazard function $\lambda_{0k}(t_D)$ for competing risks survival data can be defined based on fixed time points vectors τ_{Dk} and a baseline hazard vector $\mathbf{g}_{Dk}$.

Because the integral in model (6) cannot be evaluated analytically, we adopt adaptive Gaussian quadrature method to approximate the integral. Specifically, a quadrature method approximates a given integral by a weighted sum over predefined abscissas for the random effects u . Compared with the non-adaptive Gaussian quadrature with equally spaced points, the adaptive Gaussian quadrature method possesses the advantages of accounting for the shape of the likelihood when placing quadrature points and providing a better integration approximation²¹. Moreover, the adaptive Gaussian quadrature method can be easily implemented in SAS procedure NLMIXED because it only requires the conditional likelihood (given the random effects) and the approximated marginal likelihood can be directly maximized in SAS to obtain parameter estimates and standard errors.

3 Simulation studies

We evaluate the maximum likelihood inference of the proposed model via simulation. The simulated data structure is similar to the motivating dataset with two competing risks (death from breast cancer or other causes), strong association between IBTR and competing risks, and weak association between two competing risks. We consider two covariates with x_1 being simulated from $N(2, 2.25)$ and x_2 being simulated from $N(1, 0.25)$. We simulate 1,000 datasets with sample size $N = 1,000$ patients.

We generate the unobserved true IBTR status y from model (1) with $\alpha = (1, -1.2, 1.7)$ and $\mathbf{x}_p = (1, x_1, x_2)$. We set the baseline hazard functions for two competing risks as $\lambda_1(t) = 6.06t^{1.33}$ and $\lambda_2(t) = 3.43t^{1.44}$. We simulate the uncensored event times D_1 and D_2 from model (4), subject to independent censoring time C , with $\eta_1 = (-1, 3)$, $\eta_2 = (1, -0.2)$, $v_2 = -0.25$, $\mathbf{x}_D = (x_1, x_2)$, and random effects $u \sim N(0, 2)$. We then obtain $t_D = \min(C, D_1, D_2)$ and d . Conditional on y , we simulate relapse time t_R either from model (2) or the hazard

function model for TR patients with $r_0(t) = 11.71t^{2.19}$, $\beta = (1.6, 1.2)$, $\mathbf{x}_R = (x_1, x_2)$, and $\zeta = 3$. We set $R = 1$ and simulate z with sensitivity p and specificity q if $t_R \leq t_D$, and set $R = 0$ (IBTR is not observed) if $t_R > t_D$.

Table 1 reports the bias, the standard error (SE) of the parameter estimates, the mean of the standard error estimates (SEM), and coverage probability (CP) of the 95% confidence interval for each parameter of interest under the proposed model. In Table 1 the bias is negligible, SE is close to SEM, and the 95% confidence interval coverage rates are reasonably close to the 0.95 nominal level. The simulation results suggest that the proposed model can account for the misclassification and successfully recover the true parameters. To facilitate easy implementation of our proposed model, the SAS code has been posted in the Web Supplement.

4 Real Data Analysis

In this section, we present the results of analyzing the MDACC breast cancer dataset with the proposed method. The two competing risks are breast cancer death (251 subjects, $d = 1$) and other death (212 subjects, $d = 2$), in addition to independent censoring (4,014 subjects, $d = 0$). The covariates under consideration are age at breast cancer diagnosis (x_1), distant recurrence (developed in organs other than breasts, e.g., bones, lung, brain, liver; x_2), and tumor stage (1 if stage II or higher and 0 otherwise; x_3)²². The means and standard deviations of these variables are shown in Table 2. We divide variables x_1 , time to IBTR t_R , and time to death/censoring t_D by 10. We let $\mathbf{x}_P = (1, x_1, x_2, x_3)$ in model (1), $\mathbf{x}_R = (x_1, x_3)$ in model (2), and $\mathbf{x}_D = (x_1, x_3)$ in model (4). We use the piecewise constant baseline hazard functions in models (2) and (4) with cutpoints at every $1/M$ th quantiles with $M = 4$. We have also explored other selections of M and obtained similar results. Table 3 provides the means, SE, p values, and 95% confidence interval (CI) of the parameters from model (6).

The upper part of Table 3 shows the results of modeling the probability of IBTR being NP. The odds ratio of having NP status for the patients with distant recurrence is 0.012 (i.e., $\exp(-4.422)$; 95% CI: $[4.222e - 4, 0.342]$), compared with the patients without. The covariates age at diagnosis and tumor stage do not show statistical significance. These findings are consistent with the results of previous studies^{18,23}. The diagnostic test's sensitivity and specificity are estimated to be 0.649 (95% CI: $[0.492, 0.805]$) and 0.873 (95% CI: $[0.757, 0.989]$), respectively, indicating that 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.

The second portion of Table 3 shows the covariate effects on the hazard of IBTR. The hazard rate of IBTR for every 10-year increase in age at breast cancer diagnosis is 0.586 (i.e., $\exp(-0.535)$; 95% CI: $[0.519, 0.660]$). The hazard rate of IBTR for TR patients is 6.184 (i.e., $\exp(1.822)$; 95% CI: $[3.108, 12.293]$), compared with NP patients. The large hazard difference between TR and NP statuses is visually displayed in the large gap in the cumulative incidence curves in Figure 1 and has also been reported previously^{18,23}. The next two portions of Table 3 display the covariate effects on the competing risks of death. The

hazard rate of breast cancer death for every 10-year increase in age at breast cancer diagnosis is 0.760 (i.e., $\exp(-0.275)$; 95% CI: [0.674, 0.856]). The hazard rate for the patients with higher tumor stage is 2.971 (i.e., $\exp(1.089)$; 95% CI: [2.217, 3.979]), compared with the patients with lower tumor stage. In comparison, the hazard rate of other death for every 10-year increase in age at breast cancer diagnosis is 2.065 (i.e., $\exp(0.725)$; 95% CI: [1.828, 2.330]). The hazard rate of other death for the patients with higher tumor stage is 0.839 (i.e., $\exp(-0.176)$; 95% CI: [0.572, 1.229]), compared with the patients with lower tumor stage.

The results from the proposed model suggest that patients with earlier IBTR relapse are more likely to die from breast cancer earlier ($\hat{\gamma} = 0.534$, $p < 0.001$, 95% CI: [0.251, 0.818]). The strong association between time to IBTR and time to breast cancer death can be visualized in the large gap between curves in the left panel of Figure 2. Although the two competing risks of death are not significantly associated ($\hat{v}_2 = -0.154$, $p = 0.728$, 95% CI: [-1.022, 0.714]), patients with earlier IBTR relapse are less likely to die from other causes as suggested by the opposite signs of the estimates of γ and v_2 . This phenomena can be also visualized in the large gap between curves in the right panel of Figure 2 and the explanation is that the patient with IBTR, especially those with TR relapse, are more likely to die from breast cancer and thus less likely to die from other causes.

5 Discussion

In this article, we propose a unified framework for simultaneously modeling the binomial regression with a single misclassified binary outcome without a gold standard and the correlated time to IBTR, subject to dependent competing terminal events. The shared random effects link together the models of time to IBTR and time to the competing terminal events. One simulation study is conducted to illustrate the proposed model. Also, the motivating MDACC breast cancer dataset is analyzed as an illustrative example. Our analysis results provide useful information in identifying and quantifying the covariate effects on the probability of IBTR status and various survival times, in addition to identifying the strong association between time to IBTR and the competing risks of death. We have found that the diagnostic test are more likely to correctly classify TR patients than NP patients. Moreover, the diagnostic test based on clinical and pathological criteria has relatively low sensitivity and specificity. In the near future, researchers should develop a diagnostic test with higher accuracy using the molecular criteria¹⁸. The proposed model can be conveniently fit using adaptive Gaussian quadrature tools implemented in SAS procedure NLMIXED and can be easily implemented by other researchers. The proposed model can be broadly applied to many studies with similar data structure consisting of misclassified binary outcome and dependent survival times. For example, higher expression level of the chemokine CXCL12 (high vs. low/moderate) is an important predictor of poor survival in ovarian cancer²⁴. However, tumor grade and CXCL12 expression level can be misclassified and there is no gold standard available.

To estimate the diagnostic accuracy without a gold standard, it requires information on the misclassification structure to make the model identifiable¹¹. In model (1) for the IBTR status, the covariate vector $\mathbf{x}_P$ (i.e., age, distant recurrence, and tumor stage in the MDACC

dataset) plays the role of instrumental variables and it makes the sensitivity and specificity parameters identifiable when it has sufficient numbers of different realizations¹². In addition, the survival information included in model (6) is an important determinant of IBTR status, manifested by the clear dichotomy in the cumulative incidence curves in Figures 1 and 2. Furthermore, the simulation results (with data structure similar to the real data example) indicate that the proposed model is identifiable and the true parameters can be recovered accurately. For detailed discussion of the identifiability, modeling, and parameter estimation of multiple conditional dependent diagnostic tests, please refer to Jones *et al*²⁵.

The proposed modeling strategy has several limitations that can be addressed in future research. One limitation is the shared random effects assumption, in which two survival models (2) and (4) share a function of the random effects. While this shared random effects model is easy to implement and has low dimension of the random effects distribution, it makes a relatively strong assumption about the association between survival times. Molenberghs and Verbeke²⁶ pointed out that this kind of shared parameter model is a special case of multivariate random effects model, which in this context assumes two survival processes (i.e., times to IBTR and breast cancer death) are linked by a vector of two correlated random effects following a bivariate normal distribution. Another issue is the non-differential assumption (i.e., the sensitivity and specificity do not depend on covariates). We would like to investigate the methods to relax this assumption in the future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The project described was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant KL2 TR000370. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Fisher B, Anderson S, Bryant J, Margolese R, Deutsch M, Fisher E, et al. 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. [PubMed: 12393820]
2. Chen A, Meric-Bernstam F, Hunt K, Thames H, Oswald M, Outlaw E, et al. Breast conservation after neoadjuvant chemotherapy: the MD Anderson Cancer Center experience. *Journal of Clinical Oncology*. 2004; 22:2303–2312. [PubMed: 15197191]
3. Gustafson, P. Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments. CRC Press; 2003.
4. Albert PS. Imputation approaches for estimating diagnostic accuracy for multiple tests from partially verified designs. *Biometrics*. 2007; 63(3):947–957. [PubMed: 17825024]
5. 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.
6. 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.

7. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics*. 2001; 57(1):158–167. [PubMed: 11252592]
8. 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. [PubMed: 15180668]
9. Albert PS. Estimating diagnostic accuracy of multiple binary tests with an imperfect reference standard. *Statistics in Medicine*. 2009; 28(5):780–797. [PubMed: 19101935]
10. Zhang B, Chen Z, Albert PS. Estimating diagnostic accuracy of raters without a gold standard by exploiting a group of experts. *Biometrics*. 2012; 68(4):1294–1302. [PubMed: 23006010]
11. Ren D, Stone R. A Bayesian adjustment for covariate misclassification with correlated binary outcome data. *Journal of Applied Statistics*. 2007; 34:1019–1034.
12. 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; 7:739–744. [PubMed: 3043622]
13. 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.
14. 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): 2320–2334. [PubMed: 22996169]
15. Luo S, Su X, DeSantis S, Huang X, Yi M, Hunt K. Joint model for a diagnostic test without a gold standard in the presence of a dependent terminal event. *Statistics in Medicine*. 2014; 33(15):2554–2566. [PubMed: 24473943]
16. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 2000; 1(4):465–480. [PubMed: 12933568]
17. Elashoff RM, Li G, Li N. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*. 2008 Dec.64:762–771. [PubMed: 18162112]
18. Huang E, Buchholz T, Meric F, Krishnamurthy S, Mirza N, Ames F, et al. 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. [PubMed: 12412158]
19. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*. 1988; 16:1141–1154.
20. 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.
21. 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.
22. Singletary S, Allred C, Ashley P, Bassett L, Berry D, Bland K, et al. Revision of the American Joint Committee on Cancer staging system for breast cancer. *Journal of Clinical Oncology*. 2002; 20(17):3576–3577. [PubMed: 12202656]
23. Komoike Y, Akiyama F, Iino Y, Ikeda T, Tanaka-Akashi S, Ohsumi S, et al. 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. [PubMed: 15858440]
24. Popple A, Durrant L, Spendlove I, Rolland P, Scott I, Deen S, et al. The chemokine, CXCL12, is an independent predictor of poor survival in ovarian cancer. *British Journal of Cancer*. 2012; 106:1306–1313. [PubMed: 22415233]
25. 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. [PubMed: 19764953]
26. Molenberghs, G.; Verbeke, G. *Models for Discrete Longitudinal Data*. Springer Verlag; 2005.

Figure 1. Cumulative incidence of IBTR for the patients with IBTR status classified as either NP or TR ($R = 1$, sample size $N = 397$). P-values are from tests by Gray¹⁹. IBTR: ipsilateral breast tumor relapse. TR: true local recurrence tumor. NP: new primary tumor.

Figure 2.

Cumulative incidence of breast cancer death (left panel) and other death (right panel) of all patients (sample size $N = 4,477$). P-values are from tests by Gray¹⁹. IBTR: ipsilateral breast tumor relapse. TR: true local recurrence tumor. NP: new primary tumor.

Figure 3.

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, t_D is time from BCT to death or censoring, d is event type (0 if censor, 1 if breast cancer death, 2 if other death). BCT, breast conservation therapy; IBTR, ipsilateral breast tumor relapse.

Table 1

Simulation results of fitting the proposed model.

Parameter	Bias	SE	SEM	CP
For the probability of IBTR being NP				
$\alpha_0 = 1$	0.005	0.234	0.229	0.937
$\alpha_1 = -1.2$	-0.011	0.118	0.114	0.942
$\alpha_2 = 1.7$	0.021	0.247	0.231	0.936
$p = 0.8$	0.004	0.041	0.038	0.924
$q = 0.9$	0.001	0.017	0.016	0.929
For the time to IBTR				
$\beta_1 = 1.6$	0.006	0.093	0.095	0.953
$\beta_2 = 1.2$	-0.009	0.135	0.134	0.939
$\gamma = 0.5$	0.012	0.152	0.177	0.950
$\zeta = 3$	0.012	0.206	0.215	0.952
For the time to breast cancer death				
$\eta_{11} = -1$	-0.019	0.128	0.152	0.963
$\eta_{12} = 3$	0.044	0.369	0.433	0.951
$\sigma_u^2 = 2$	0.127	0.820	1.057	0.931
For the time to other death				
$\eta_{21} = 1$	0.008	0.052	0.053	0.956
$\eta_{22} = -0.2$	0.002	0.095	0.095	0.951
$\nu_2 = -0.25$	-0.010	0.086	0.091	0.950

Table 2

Characteristics of some variables in the MDACC breast cancer dataset

Covariates	mean (sd)
Age at diagnosis (yr)	55.4 (12.1)
Distant recurrence (%)	7.5
Stage (%)	23.1

Table 3

The maximum likelihood estimates (MLE), standard errors (SE), p-values, and 95% confidence interval (CI) of the parameters from model (6).

	MLE	SE	p	95% CI
For the probability of IBTR being NP				
intercept	4.282	2.533	0.091	−0.683, 9.248
Age	−0.056	0.309	0.856	−0.661, 0.549
Distant	−4.422	1.708	0.010	−7.770, −1.074
Stage	−0.726	0.900	0.420	−2.490, 1.038
p	0.649	0.080		0.492, 0.805
q	0.873	0.059		0.757, 0.989
For the time to relapse				
Age	−0.535	0.061	< 0.000	−0.656, −0.415
Stage	−0.014	0.170	0.933	−0.349, 0.320
γ	0.534	0.144	< 0.001	0.251, 0.818
ζ	1.822	0.351	< 0.001	1.134, 2.509
For the time to breast cancer death				
Age	−0.275	0.061	< 0.001	−0.394, −0.155
Stage	1.089	0.149	< 0.001	0.796, 1.381
σ_u^2	1.592	0.434	< 0.001	0.742, 2.442
For the time to other death				
Age	0.725	0.062	< 0.001	0.603, 0.846
Stage	−0.176	0.195	0.365	−0.558, 0.206
v_2	−0.154	0.443	0.728	−1.022, 0.714