

Modeling intra-tumor protein expression heterogeneity in tissue microarray experiments

Ronglai Shen^{1,*}, Debashis Ghosh² and Jeremy M. G. Taylor³

¹*Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, U.S.A.*

²*Department of Statistics and Huck Institute of Life Sciences, Penn State University, University Park, PA, U.S.A.*

³*Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.*

SUMMARY

Tissue microarrays (TMAs) measure tumor-specific protein expression via high-density immunohistochemical staining assays. They provide a proteomic platform for validating cancer biomarkers emerging from large-scale DNA microarray studies. Repeated observations within each tumor result in substantial biological and experimental variability. This variability is usually ignored when associating the TMA expression data with patient survival outcome. It generates biased estimates of hazard ratio in proportional hazards models. We propose a Latent Expression Index (LEI) as a surrogate protein expression estimate in a two-stage analysis. Several estimators of LEI are compared: an empirical Bayes, a full Bayes, and a varying replicate number estimator. In addition, we jointly model survival and TMA expression data via a shared random effects model. Bayesian estimation is carried out using a Markov chain Monte Carlo method. Simulation studies were conducted to compare the two-stage methods and the joint analysis in estimating the Cox regression coefficient. We show that the two-stage methods reduce bias relative to the naive approach, but still lead to under-estimated hazard ratios. The joint model consistently outperforms the two-stage methods in terms of both bias and coverage property in various simulation scenarios. In case studies using prostate cancer TMA data sets, the two-stage methods yield a good approximation in one data set whereas an insufficient one in the other. A general advice is to use the joint model inference whenever results differ between the two-stage methods and the joint analysis. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: biomarker; empirical Bayes; joint modeling; mixed effects; tissue microarray; varying replicate number (VRN)

*Correspondence to: Ronglai Shen, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 307 East 63rd Street, 3rd Floor, New York, NY 10065, U.S.A.

†E-mail: shenr@mskcc.org

Contract/grant sponsor: NIH

1. INTRODUCTION

DNA microarray technology has enabled expression measurement of thousands of genes simultaneously, providing a platform for rapid screening of genomic biomarkers in cancer. Translating these discovery-type findings into clinical relevance is a more challenging task. Gene expression profiling studies using spotted cDNA arrays or Affymetrix GeneChip arrays often yield a few hundred candidate cancer genes that are associated with a phenotype. Only a small portion of these will be eventually validated for the corresponding expression changes at the protein level. The advent of tissue microarray (TMA) technology has provided a proteomic platform for such validation studies to find clinically useful biomarkers. TMA experiments measure tumor-specific protein expression via high-density immunohistochemical staining assays, allowing simultaneous evaluation of hundreds of patient samples in a single experiment [1]. Since their initial development, TMA-based expression studies have quickly become an integral part of cancer biomarker development [2–4].

A typical tissue array slide comprises up to 1000 tiny biopsy tissue elements, which we will refer to as cores, with multiple cores corresponding to repeated sampling from the same tumor. Expression measures on these replicate cores constitute the TMA core-level data. These can display substantial within-subject variability for both biological and experimental reasons. Biologically, for tumors that are highly infiltrative and heterogeneous in nature (e.g. prostate tumors), protein expression pattern can be quite variable. For example, cell proliferation genes often exhibit localized high expression within a tumor, indicating elevated aggressiveness and metastatic potential in the corresponding areas. Replicate sampling from various regions of the tumor is therefore important in capturing the underlying heterogeneity within a tumor. Experimental sources of the variability can come from a combination of probe affinity, measurement imprecision, and further missing data due to insufficient sampling. Without appropriately accounting for these variabilities, the noise-prone measurements tend to attenuate the prognostic value of a potential biomarker in predicting disease outcome. The lack of a model-based approach for TMA core-level expression data to effectively model the intra-tumor variation has motivated us to carry out a full investigation.

A good analogy for understanding TMA data structure is from probe-level data generated by the Affymetrix GeneChip arrays. GeneChip arrays measure gene expression at the mRNA transcripts level. The probe-level data refer to the replicate expression measures on a set of 16–20 small oligonucleotide probe sets derived for a target gene. The biological variation comes primarily from these oligonucleotide probe sequence variants, whereas experimental variations arise during the process of slide printing, hybridization, and optical reading. Li and Wang [5] reported that the variation of a specific probe across multiple arrays could be considerably smaller than the variance across probes within a probe set. Modeling Affymetrix probe-level data has generated much attention [5, 6] as the technology has become more mature and widely used.

Similarly, in TMA experiments, modeling within-tumor protein expression heterogeneity is an important problem. In these tissue-based experiments, the variation across core samples within a tumor can be substantially larger than the variation observed across subjects. Etzioni *et al.* [7] used a compositional analysis to model such heterogeneity and compared the proportion of cells stained at different intensity levels between normal and tumorous tissues. In this study, we focus on the effect of modeling intra-tumor variation in the context of predicting patient survival outcome. In a latent variable modeling framework, we assume that an underlying ‘true’ expression value

predicts survival. In real experiments, this ‘true’ expression cannot be precisely measured due to sampling variabilities and measurement imprecision. Instead, we observe the core-level expression measurements that are subject to these measurement errors. In a two-stage method, we adapt ideas from measurement error modeling and propose a Latent Expression Index (LEI) to approximate the underlying true value and focus on its behavior in proportional hazards models. We adapt an empirical Bayes (EB) estimator [8] to (1) incorporate important clinical parameters such as Gleason score and pathological stage of the tumor and (2) adjust for the varying number of cores. We further establish a joint model for TMA core-level data and survival outcome via a shared random effect. There is a large literature on joint modeling of longitudinal data and survival [9–16]. These methods have been developed predominantly for modeling survival and CD4 counts in AIDS patients; here their application to TMA data in cancer biomarker studies is novel. Using both simulations and two published TMA data sets, we compare the performances of the naive, two-stage LEI and the joint model approach in obtaining parameter estimates and associated inference.

The article is organized as follows. Section 2 specifies notation and models for the TMA core-level expression data and for the patient survival data. Section 3 introduces LEI and its use in a two-stage method. Section 4 presents the joint modeling approach and the Bayesian estimation framework. Simulation results to compare the performances of these methods are then discussed in Section 5. Case studies using two prostate cancer TMA data sets are presented in Section 6. Further discussion can be found in Section 7.

2. MODEL SPECIFICATION

Measurement model for the TMA core-level data

Let X_i^* be the latent expression value for a biomarker in tumor i , $i = 1, \dots, n$. Assume that the observed TMA core-level measurement for the j th core in the i th tumor is

$$X_{ij} = X_i^* + U_{ij}, \quad j = 1, \dots, r_i, \quad i = 1, \dots, n, \quad (1)$$

where we assume $X_i^* \sim N(\mu_{x^*}, \sigma_{x^*}^2)$. The mean μ_{x^*} is a linear function of clinical covariates: $\mu_{x^*} = \theta_0 + \boldsymbol{\theta} \mathbf{Z}_i'$, where $\mathbf{Z}_i = (Z_{1i}, Z_{2i}, \dots, Z_{pi})$ constitutes a row vector of p clinical parameters characterizing histologic and pathologic features of the tumor, and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ is the associated p -dimensional row vector of effect sizes. In this model, U_{ij} represents the variation of the r_i core-level expression measurements about the ‘true’ value X_i^* . We assume that U_{ij} is i.i.d. $N(0, \sigma_u^2)$ and independent of X_i^* .

Survival model

Let the observed survival time for patient i ($i = 1, \dots, n$) be $T_i = \min(Y_i, C_i)$, where Y_i is the time from diagnosis to disease recurrence; C_i is the time to censoring which is independent of Y_i , and $\delta_i = 1\{Y_i < C_i\}$ is the censoring indicator. Under the Cox proportional hazards model, the hazard rate for patient i is

$$\lambda(t) = \lambda_0(t) e^{\beta^* X_i^* + \mathbf{b} \mathbf{Z}_i'} \quad (2)$$

where $\lambda_0(t)$ is the baseline hazard function, β^* is the regression coefficient associated with the ‘true’ protein expression, and $\mathbf{b} = (b_1, \dots, b_p)$ is the row vector of coefficients associated with the

vector of clinical variables \mathbf{Z}'_i . We also consider a parametric Weibull regression model with the following form for the hazard function:

$$\lambda(t) = \gamma t^{\gamma-1} e^{\beta_0 + \beta^* X_i^* + \mathbf{bZ}'_i} \tag{3}$$

3. TWO-STAGE PLUG-IN METHOD

Given the rare disease assumption and the assumption that the measurement error U_{ij} has no predictive value, i.e. $\lambda(t|\mathbf{X}_i, X_i^*) = \lambda(t|X_i^*)$, Prentice [17] introduced the induced hazard rate function for Cox regression model with covariates measured with error. When normality is assumed for the error distribution, a first-order bias-correction method is to replace X_i^* with its conditional mean

$$\lambda(t|\mathbf{X}_i) = \lambda_0(t) e^{\beta^* E[X_i^*|\mathbf{X}_i, \mathbf{Z}_i] + \mathbf{bZ}'_i} \tag{4}$$

and estimate β^* by maximizing the corresponding partial likelihood. Here $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ir_i})$ denotes the vector of observed expression measures for tumor i . Note that the regression calibration model in (4) is an approximation to the true model in (2) due to the rare disease assumption and the omission of a higher-order term implicit in (2). Details of such first-order bias-correction methods can be found in [17, 18]. Define the LEI to be an estimate of the conditional mean, $LEI_i = \hat{E}[X_i^*|\mathbf{X}_i, \mathbf{Z}_i]$, for each subject i . A two-stage plug-in method can be described by the following algorithm: (1) compute LEI_i ($i = 1, \dots, n$) as a surrogate expression estimate that adjusts for measurement error and (2) apply the Cox or Weibull regression model using LEI_i to obtain an estimate of β^* and the associated standard error.

In the following section, we describe methods for computing LEI for TMA data. These include an EB estimator conditional on clinical covariates, a full Bayes (FB) approach and a varying replicate number (VRN) method as an extension to adjust for the number of cores per tumor.

3.1. Methods for computing LEI

3.1.1. EB and FB estimators. Express (1) as a mixed effects model

$$X_{ij} = \theta_0 + \boldsymbol{\theta}\mathbf{Z}'_i + v_i + U_{ij} \tag{5}$$

where $v_i \sim N(0, \sigma_{x^*}^2)$. To make a connection between (5) and (1), X_i^* is the unobserved ‘true’ protein expression level, which can be expressed as $X_i^* = \theta_0 + \boldsymbol{\theta}\mathbf{Z}'_i + v_i$. Subsequently, v_i , which is normally distributed with mean zero and variance, has the interpretation of a mean centered ‘true’ protein expression level. The EB estimator can then be derived as

$$LEI_i^{eb} = \hat{\gamma}_i \bar{X}_i + (1 - \hat{\gamma}_i)(\hat{\theta}_0 + \hat{\boldsymbol{\theta}}\mathbf{Z}'_i) \tag{6}$$

where $\hat{\gamma}_i \equiv \hat{\sigma}_{x^*}^2 (\hat{\sigma}_{x^*}^2 + \hat{\sigma}_u^2 r_i^{-1})^{-1}$ is the attenuation factor [18]. Parameter estimates $\{\hat{\theta}_0, \hat{\boldsymbol{\theta}}, \hat{\sigma}_u^2, \hat{\sigma}_{x^*}^2\}$ can be obtained by fitting a mixed effects model as described in (5), using a restricted maximum likelihood approach [19, 20].

The EB estimator conditions on the set of parameter estimates derived from the data. The uncertainty of these estimates are not accounted for in LEI^{eb} . For this reason, we also consider an

FB estimator, LEI_i^{fb} , where standard conjugate hyperprior distributions are adopted: $\sigma_u^{-2}, \sigma_{x^*}^{-2} \sim \Gamma(r_0, \gamma_0)$. The FB estimator

$$LEI_i^{fb} = \tilde{\theta}_0 + \tilde{\theta}'Z'_i + \tilde{v}_i$$

is then based on the posterior inference from model (5) where $\{\tilde{\theta}_0, \tilde{\theta}, \tilde{v}_i\}$ are the posterior means given the data.

3.1.2. The VRN method. In a typical TMA construction, K_i cores are placed on the array for each tumor i . However, not all of the measurements $\{X_{i1}, \dots, X_{iK_i}\}$ are available. Several reasons contribute to a varying number of replicate measure. These include heterogeneous tissue composition and technical defects such as image corruption. The expression measurement from non-tumorous tissue types or a corrupted image is typically considered unsuitable for an outcome analysis and excluded. Let $M_{ij} = 0, j = 1, \dots, K_i$ indicate that the j th core from the i th tumor is lost due to the aforementioned reasons and $M_{ij} = 1$ if it is available. Expression measures are retained for $r_i \equiv \sum_{j=1}^{K_i} M_{ij}$ cores, where r_i varies across tumor samples and possibly depends on covariate Z_i . We consider r_i to follow a binomial distribution given K_i and $P(M_{ij} = 1)$ with possible over-dispersion. Assuming independence and homogeneity across j for M_{ij} , we adopt a logistic mixed effects model:

$$\text{logit}P(M_{ij} = 1) = \psi_{0i} + \Psi Z'_i \tag{7}$$

where $\psi_{0i} \sim N(\psi_0, \sigma_\psi^2)$, $Z_i = (Z_{1i}, Z_{2i}, \dots, Z_{gi})$ is the vector of g clinical covariates that can be the same or different from those in (5), and $\Psi = (\psi_1, \psi_2, \dots, \psi_g)$ is the associated vector of coefficients. Therefore

$$r_i \sim \text{Binomial} \left(K_i, p_i = \frac{e^{\psi_{0i} + \Psi Z'_i}}{1 + e^{\psi_{0i} + \Psi Z'_i}} \right) \tag{8}$$

The expression index under the VRN model is then derived by averaging over all the possible values of (r_i, K_i) . In particular,

$$\begin{aligned} LEI_i^{vm} &= E_{(r_i, K_i)} E[X_i^* | \mathbf{X}_i, Z_i, r_i, K_i] \\ &= \sum_{s=1}^R \sum_{m=0}^s \left\{ \frac{\hat{\sigma}_{x^*}^2 m}{\hat{\sigma}_{x^*}^2 m + \hat{\sigma}_u^2} \bar{X}_i + \frac{\hat{\sigma}_u^2}{\hat{\sigma}_{x^*}^2 m + \hat{\sigma}_u^2} (\hat{\theta}_0 + \hat{\theta}'Z'_i) \right\} \\ &\quad \times \binom{s}{m} \left(\frac{e^{\hat{\psi}_{0i} + \hat{\Psi}Z'_i}}{1 + e^{\hat{\psi}_{0i} + \hat{\Psi}Z'_i}} \right)^m \left(\frac{1}{1 + e^{\hat{\psi}_{0i} + \hat{\Psi}Z'_i}} \right)^{s-m} \hat{P}(K_i = s) \end{aligned} \tag{9}$$

Here R is taken to be $\max_i K_i$. An additional assumption for the above is that the expression measures do not correlate with r_i or K_i . Parameter estimates $\{\hat{\theta}_0, \hat{\theta}, \hat{\sigma}_u^2, \hat{\sigma}_{x^*}^2\}$ can be obtained by fitting a mixed effects model as described in (5). A logistic mixed effects model in the form

of (7) was fitted to obtain $\{\hat{\psi}_{0i}, \hat{\Psi}\}$. Estimation is via methods described in [21, 22]. The empirical proportions were used for $\hat{P}(K_i = s)$.

In a relatively balanced TMA array where the number of replicate measures r_i does not vary much across subjects, $\gamma_i \equiv \sigma_{x^*}^2 (\sigma_{x^*}^2 + \sigma_u^2 r_i^{-1})^{-1}$ is an approximately constant adjustment factor. The amount of shrinkage in LEI^{eb} toward the overall mean depends primarily on the ratio of the within- to between-subject variation in that particular data set. In our example, however, r_i is a highly variable quantity. It exerts a larger role in determining how much weight LEI_i^{eb} gives to a particular subject's data relative to the estimated population mean. The motivation for LEI^{vm} is to provide a replicate number-averaged expression estimate that alleviates the variability induced by r_i in the EB estimator.

4. JOINT MODELING OF SURVIVAL AND TMA CORE-LEVEL DATA

The two-stage approaches described above are attractive for their simplicity and straightforward interpretation. They require minimal computation and can be easily implemented using existing statistical packages. However, there are major limitations for the two-stage method [23]. First, the two-stage method involves a first-order approximation and ignores the second-order term $\beta^{*2} \sigma^2(X_i^* | \mathbf{X})$ in the induced hazard rate function (4). As we will illustrate in the simulation study, such an approximation works well when β^* is close to zero, but otherwise leads to sizeable bias in $\hat{\beta}^*$. Second, parameter estimates in the second stage do not account for the uncertainty in estimating LEI in the first stage. The associated standard error for $\hat{\beta}^*$ will be over-optimistic. Given these considerations, it is desirable to make inference based on the joint likelihood of the failure time and TMA expression data. In this article, we adopt a shared random effect model to induce correlation between the TMA data and the survival outcome.

Given the measurement model specified in (5) for the TMA data, we express the proportional hazards model for the survival outcome as

$$\lambda(t) = \lambda_0(t) \exp(\mathbf{bZ}'_i + \beta^* v_i) \tag{10}$$

The parameter v_i constitutes a shared random effect that connects the measurement model (5) and the survival outcome model (10). The expression data and survival times are then assumed to be independent given v_i . The joint likelihood of the observed data $\{T_i, \delta_i, \mathbf{X}_i\}$ and the random effects v_i is therefore

$$L_{\text{Joint}} = \prod_{i=1}^n f(t_i, \delta_i | v_i, \mathbf{z}_i) \times \prod_{i=1}^n f(\mathbf{x}_i | v_i, \mathbf{z}_i) f(v_i) = L_{\text{SURV}} \times L_{\text{ME}} \tag{11}$$

where

$$L_{\text{ME}} = \prod_{i=1}^n \left(\prod_{j=1}^{r_i} \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp \left\{ -\frac{(x_{ij} - \theta_0 - \boldsymbol{\theta}\mathbf{Z}'_i - v_i)^2}{2\sigma_u^2} \right\} \right) \frac{1}{\sqrt{2\pi\sigma_{x^*}^2}} \exp \left\{ -\frac{v_i^2}{2\sigma_{x^*}^2} \right\} \tag{12}$$

$$L_{\text{SURV}} = \prod_{i=1}^n \prod_{l=1}^L \lambda_l^{d_l} \exp \left(\sum_{i \in D_l} \mathbf{bZ}'_i + \beta^* v_i \right) \exp \left(-\lambda_l \sum_{i \in R_l} \Delta_{il} e^{\mathbf{bZ}'_i + \beta^* v_i} \right)$$

We used a piecewise constant hazards model in which the time axis is partitioned into L disjoint intervals, I_1, \dots, I_L , where $I_l = [a_{l-1}, a_l)$ with $a_0 < t_i$ and $a_L > t_i$ for all $i = 1, \dots, n$. Assume a constant baseline hazard in the l th interval, $\lambda_0(t) = \lambda_l$ for $t \in I_l$. R_l is the set at risk at the beginning of interval l ; d_l is the number of failures in interval l ; and $\Delta_{il} = \min(t_i, a_l) - a_{l-1}$.

Alternatively, a parametric Weibull model can be assumed for the survival outcome using the following hazard function:

$$\lambda(t) = \gamma t^{\gamma-1} \exp(b_0 + \mathbf{bZ}'_i + \beta^* v_i) \quad (13)$$

When $\gamma = 1$, the above reduces to exponential distribution with constant failure rate $\exp(b_0 + \mathbf{bZ}'_i + \beta^* v_i)$. The survival time component of the joint likelihood in (12) is then replaced by

$$L_{\text{SURV}} = \prod_{i=1}^n \{\gamma t_i^{\gamma-1} e^{\gamma(b_0 + \mathbf{bZ}'_i + \beta^* v_i)}\}^{\delta_i} \exp(-e^{\gamma(b_0 + \mathbf{bZ}'_i + \beta^* v_i)} t_i^\gamma) \quad (14)$$

In a Bayesian estimation framework, the following prior distributions are specified for the model parameters:

$$\begin{aligned} (\beta^*, \theta_0, \theta, b_0, b) &\sim N(\mu_0, \sigma_0^2) \\ (\sigma_u^{-2}, \sigma_{x^*}^{-2}) &\sim \Gamma(r_0, \gamma_0) \\ (\gamma, \lambda_l, l = 1, \dots, L) &\sim \Gamma(r_0, \gamma_0) \end{aligned} \quad (15)$$

The prior choices in (15) render proper noninformative prior distributions with the hyperparameters chosen to achieve large variances (wide spread). In particular, $\mu_0 = 0$, $\sigma_0^2 = 10000$, $r_0 = 0.001$, $\gamma_0 = 0.001$. The values of r_0 and γ_0 are set to be small to impose noninformativeness (although there have been criticisms that the resulting posterior distributions remain highly sensitive to the choice of r_0 and γ_0). We had tested different sets of values for r_0 and γ_0 : (0.1, 0.1), (0.01, 0.01), and (0.001, 0.001). When the sample size is moderate to large, the variance estimates, $\hat{\sigma}_u$ and $\hat{\sigma}_{x^*}$, are well behaved and not significantly affected by the choice of r_0 and γ_0 . Samples from the posterior distribution are obtained using Markov chain Monte Carlo (MCMC) methods. In the likelihood function of (11), the Bayesian estimation procedure treats v_i as missing data and imputes it at each MCMC iteration via posterior sampling from the conditional distribution of v_i given the rest of the variables. One then proceeds with the likelihood function in (11) as if v_i were observed.

5. SIMULATION

5.1. Simulation setup

The additive measurement error model in (5) with one covariate Z_{1i} is used to simulate the expression measure X_{ij} , $i = 1, \dots, n$ and $j = 1, \dots, r_i$. In this model, $\theta_0 = 0$ and $\theta_1 = 1$. Furthermore, $v_i \sim N(0, 1)$, $U_{ij} \sim N(0, 0.5)$. The covariate Z_{1i} is simulated from a $N(0, 1)$ distribution. The total number of cores sampled, K_i , takes values in $\{1, 2, \dots, 12\}$ with $P(K_i = 6) = 0.4$, $P(K_i = 1) = \dots = P(K_i = 5) = 0.1$, and $P(K_i = 7) = \dots = P(K_i = 12) = \frac{1}{60}$, mimicking the proportions from the actual tissue array data set used in this study. We simulate the number of repeated measures $r_i \equiv \sum_{j=1}^{K_i} M_{ij}$ from a Binomial(K_i, p_i), where $p_i = 1 - \pi^{1/K_i}$ such that the missing proportion

equals π . The survival time T_i is simulated from a proportional hazards model in the form of (2) with $\lambda_0(t) \equiv 1$ and $\beta^* = 1$ or 2. An additional covariate Z_{1i} is further assumed to associate with T_i with the coefficient being one. The censoring time is simulated from an independent exponential distribution that results in a 30 per cent censoring proportion. Results are summarized over 100 such simulated data sets each of a sample size $n = 200$. In general, we assigned parameter values in the simulation to mimic those for the real data sets.

Computation of LEI^{eb} , LEI^{vm} were carried out using the PROC MIXED and the IML procedure in SAS (SAS Institute, Cary, NC). LEI^{fb} and the joint models were implemented using OpenBUGS via the R interface BRugs [24, 25]. We ran two chains with 1000 burn-in and 1000 updates per chain for the MCMC convergence.

5.2. Simulation results

The simulation results are summarized in Table I. For $\beta^* = 1$ in the survival models, the naive approach (using \bar{X}_i as a surrogate expression) attenuates the true effect size by around 25 per cent. The coverage probability of a nominal 95 per cent confidence interval of $\hat{\beta}^*$ is 0.10 at best. The two-stage methods (LEI) achieved a considerable bias correction by adjusting for the measurement error in the LEI imputation. The joint modeling approach gives the best estimate $\hat{\beta}^* = 1.03$ and a coverage probability of 95 per cent compared with the truth.

Next a larger effect size is simulated ($\beta^* = 2$). The bias in the two-stage approaches due to the first-order approximation is evident. Overall the two-stage methods generate less biased $\hat{\beta}^*$'s compared to the naive estimate. Nevertheless, these are 15–25 per cent smaller than the true β^* . The coverages are poor for the two-stage approaches. The joint modeling approach should be advocated in this scenario for inference. Note that the joint model estimates of β^* are slightly greater than 2, especially under the Weibull distribution. This may be driven by the prior distributions adopted for the parameters under the Bayesian estimation scheme. We have observed that such a difference disappears when the sample size gets larger.

Table I. Simulation study. Results are summarized over 100 simulated data sets each of $n = 200$.

		Weibull				Proportional hazards			
		$\hat{\beta}^*$	$s.e.(\hat{\beta}^*)$	$s.d.(\hat{\beta}^*)$	Coverage	$\hat{\beta}^*$	$s.e.(\hat{\beta}^*)$	$s.d.(\hat{\beta}^*)$	Coverage
$\beta^* = 1$	X^*	1.01	0.08	0.08	0.96	1.01	0.08	0.08	0.97
	Naive	0.75	0.07	0.08	0.10	0.75	0.07	0.08	0.09
	LEI^{eb}	0.93	0.08	0.09	0.85	0.93	0.08	0.09	0.87
	LEI^{fb}	0.93	0.08	0.09	0.85	0.93	0.08	0.09	0.88
	LEI^{vm}	0.97	0.09	0.10	0.89	0.96	0.09	0.10	0.87
	Joint model	1.03	0.11	0.11	0.95	1.03	0.11	0.11	0.95
$\beta^* = 2$	X^*	2.05	0.12	0.12	0.95	2.03	0.11	0.10	0.94
	Naive	1.13	0.08	0.10	0	1.18	0.08	0.11	0
	LEI^{eb}	1.51	0.11	0.13	0.03	1.58	0.11	0.13	0.06
	LEI^{fb}	1.48	0.11	0.12	0	1.55	0.10	0.12	0.04
	LEI^{vm}	1.62	0.11	0.14	0.19	1.70	0.11	0.15	0.30
	Joint model	2.16	0.27	0.30	0.89	2.07	0.19	0.20	0.93

6. CASE STUDY IN PROSTATE CANCER

6.1. Data description

In this study, we consider two prostate tumor TMA data sets. The α -methylacyl CoA racemase (AMACR) is a peroxisomal and mitochondrial enzyme that plays an important role in fatty acid metabolism. AMACR has been shown to consistently overexpress in prostate tumors [26]. Rubin *et al.* [4] profiled AMACR protein expression using a TMA constructed on 203 prostate tumors from a surgical cohort who underwent radical prostatectomy at the University of Michigan as a primary therapy for clinically localized prostate cancer diagnosed between 1994 and 1998. They found that AMACR is a significant predictor of the prostate-specific antigen (PSA) failure in these 203 patients.

The second biomarker evaluated in this study is BM28. BM28 encodes a highly conserved mini-chromosome maintenance protein (MCM) that is involved in the initiation of genome replication. Bismar *et al.* [27] profiled a total of 41 genes (including BM28) in a TMA-based proteomic study. They identified that a 12-gene model showing the expression combination of the 12 genes significantly associates with tumor progression and PSA failure in a set of 79 men following surgery for clinically localized prostate cancer. The expression level of BM28 however did not show significant prognostic value in their analysis. We chose BM28 to evaluate the possibility of it being a false negative biomarker due to measurement error.

For the AMACR data, an average of $K_i = 5.5$ (range: 2–12) tissue core specimens were taken from each tumor sample and placed on a tissue array for immunohistochemical staining. After initial diagnostic evaluation of each core, an average of 29 per cent (range: 0–86 per cent) of the cores were excluded due to reasons discussed earlier, leading to 5.5 per cent missing subjects. The BM28 data have a similar tissue array design. A quantitative imaging analysis of the staining intensity was obtained using the ACIS II (Chromavision, San Juan Capistrano, CA) system. The intensity level ranges from 0 to 255 chromogen intensity units and is transformed using the natural logarithm (one unit added to avoid taking logarithm of 0) and normalized to have mean zero and standard deviation one. Disease recurrence is defined as a serum PSA increase >0.2 ng/mL after radical prostatectomy. Censored observations are those free of the recurrence at the time of last follow-up.

6.2. Measurement error and regression attenuation in TMA data analysis

Figure 1 illustrates a considerable amount of measurement error in the core-level expression data from the two TMA experiments. In the AMACR data, methods-of-moments estimates of the variance components are $\hat{\sigma}_u^2 = 0.46$ and $\hat{\sigma}_{x^*}^2 = 0.54$. In the BM28 data, the methods-of-moments estimates of the variance components are $\hat{\sigma}_u^2 = 0.62$ and $\hat{\sigma}_{x^*}^2 = 0.21$. The within-subject variation is almost three times the between-subject variation.

In a Cox proportional hazards model context, we did a simple simulation where $X_i^* \sim N(0, 1)$, $\beta^* = 1$. A naive estimator $\bar{X}_i = r_i^{-1} \sum_{j=1}^{r_i} X_{ij}$ —the average core-level expression for tumor i —is used as the surrogate expression to replace X_i^* in (2). We simulate situations where measurement error is small ($\sigma_u^2 = 0.1$), moderate ($\sigma_u^2 = 0.5$), and large ($\sigma_u^2 = 1$). Figure 2 shows various degrees of regression attenuation in the estimate of β^* as a function of replicate number r_i and the amount of error σ_u^2 . When the parameter values are set to resemble the AMACR data, the naive estimate of β^* is approximately 30 per cent smaller than the true value. With the current TMA construction

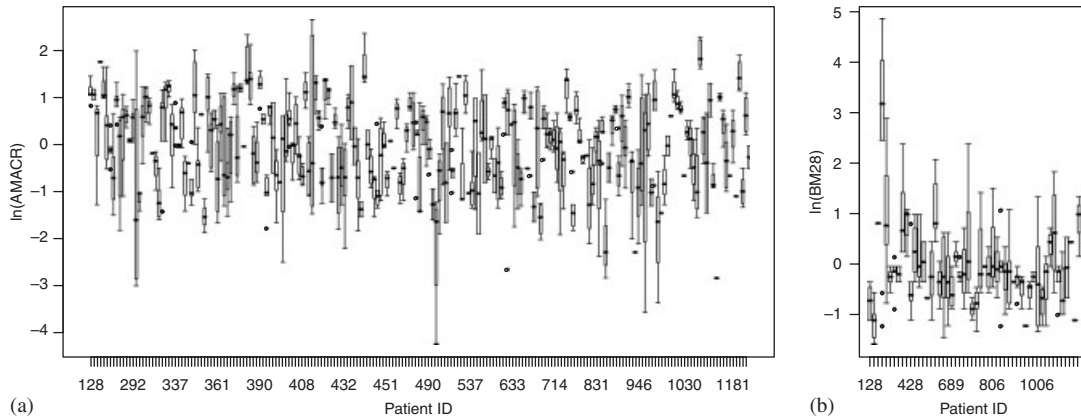


Figure 1. Variance plots to represent the within-subject variation in the TMA core-level expression data. (a) The AMACR data: estimates of the variance components are $\hat{\sigma}_u^2=0.46$ and $\hat{\sigma}_{x^*}^2=0.54$ and (b) the BM28 data: estimates of the variance components are $\hat{\sigma}_u^2=0.62$ and $\hat{\sigma}_{x^*}^2=0.21$.

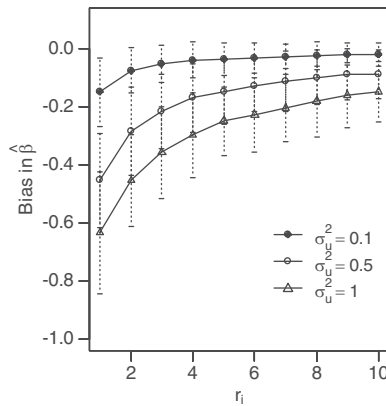


Figure 2. A simulation demonstration of the bias in Cox regression coefficient estimate as a function of the number of repeated measures r_i . The average bias with a 95 per cent CI over 100 simulated data sets of sample size $n=200$ is plotted.

protocol specifying three cores per subject due to economic and tissue-preservation reasons, and a great amount of within-subject variability routinely observed in the core-level expression data, Figure 2 effectively conveys the importance of modeling measurement error in TMA data. In the following two sections, we implement the measurement error models to demonstrate how statistical inference differs from the previous results.

6.3. AMACR expression and biochemical recurrence in prostate cancer

In prostate cancer, Gleason score, pathologic stage, and tumor size are among the most important clinical parameters. We include these as clinical covariates Z_i to adjust in the measurement

model (5), the replicate number model (7), and the survival outcome model (10). In the measurement model, $\hat{\theta}_{\text{TumorSize}} = 0.32$ with an associated standard error of 0.13, indicating a marginal association of tumor size with AMACR expression level. In the replicate number model, $\hat{\psi}_{\text{TumorSize}} = 0.72$ with an associated standard error of 0.16, which is consistent with our expectation that a larger tumor sample provides more abundant number of cores.

Table II lists the estimates and associated standard errors (posterior standard deviation) of β^* in the outcome model. The measurement error adjustment has significantly improved upon the naive estimate. The error-adjusted $\hat{\beta}^*$ is around 0.75 ($\widehat{\text{s.e.}}(\hat{\beta}^*) = 0.26$), approximately 31 per cent larger in absolute value than the naive estimate, which is 0.57 ($\widehat{\text{s.e.}}(\hat{\beta}^*) = 0.20$). The amount of attenuation in β^* is quite consistent with what we conclude from the simulated data sets in the previous section. In this data set, the two-stage methods (LEI^{eb}, LEI^{fb}, and LEI^{vm}) perform equally well as the joint modeling approach. The simplicity and computational efficiency of LEI serves as a satisfactory core-level expression index for AMACR. However, as mentioned earlier, the two-stage methods are based on a first-order approximation, the accuracy of which is largely driven by the size of β^* and the ratio of within-subject variation to between-subject variation. As will be shown in the other data example, the two-stage methods will not be always a suitable approach.

Kaplan–Meier curves are useful as a graphical representation of the prognostic value of a biomarker. We examined these plots by dividing the subjects into different risk groups based on the values of AMACR expression estimates derived under each method. In Figure 3(a), subjects with AMACR high, median, and low expression groups based on LEI^{vm} and the joint model estimates (C and D, respectively) are significantly better separated in terms of the probability of recurrence-free survival when compared with that using the naive mean estimates (A).

Table II. A case study using prostate cancer TMA data sets.

		Weibull		Proportional hazards	
		$\hat{\beta}^*$	$\widehat{\text{s.e.}}(\hat{\beta}^*)$	$\hat{\beta}^*$	$\widehat{\text{s.e.}}(\hat{\beta}^*)$
AMACR ($n = 203$)	Naive	-0.573	0.201	-0.571	0.198
	LEI ^{eb}	-0.761	0.266	-0.751	0.266
	LEI ^{fb} *	-0.753	0.262	-0.752	0.263
	LEI ^{vm}	-0.735	0.260	-0.737	0.258
	Joint modeling*	-0.742	0.268	-0.745 [†]	0.268
BM28 ($n = 52$)	Naive	0.828	0.414	0.900	0.399
	LEI ^{eb}	1.051	0.519	1.146	0.494
	LEI ^{fb} [†]	1.457	0.564	1.493	0.504
	LEI ^{vm}	1.034	0.552	1.151	0.537
	Joint modeling*	1.753	0.766	1.592 [†]	0.600

*Estimation is based on MCMC methods, where we sampled two chains each with 10 000 burn-in and 10 000 updates.

[†]We used $J = 5$ and 3 intervals for the piecewise exponential distribution in the AMACR and BM28 data set, respectively. The estimate of the Weibull shape parameter is $\hat{\gamma} = 0.71$, and $\hat{\gamma} = 0.97$ from the joint modeling for the AMACR and BM28 data set, respectively.

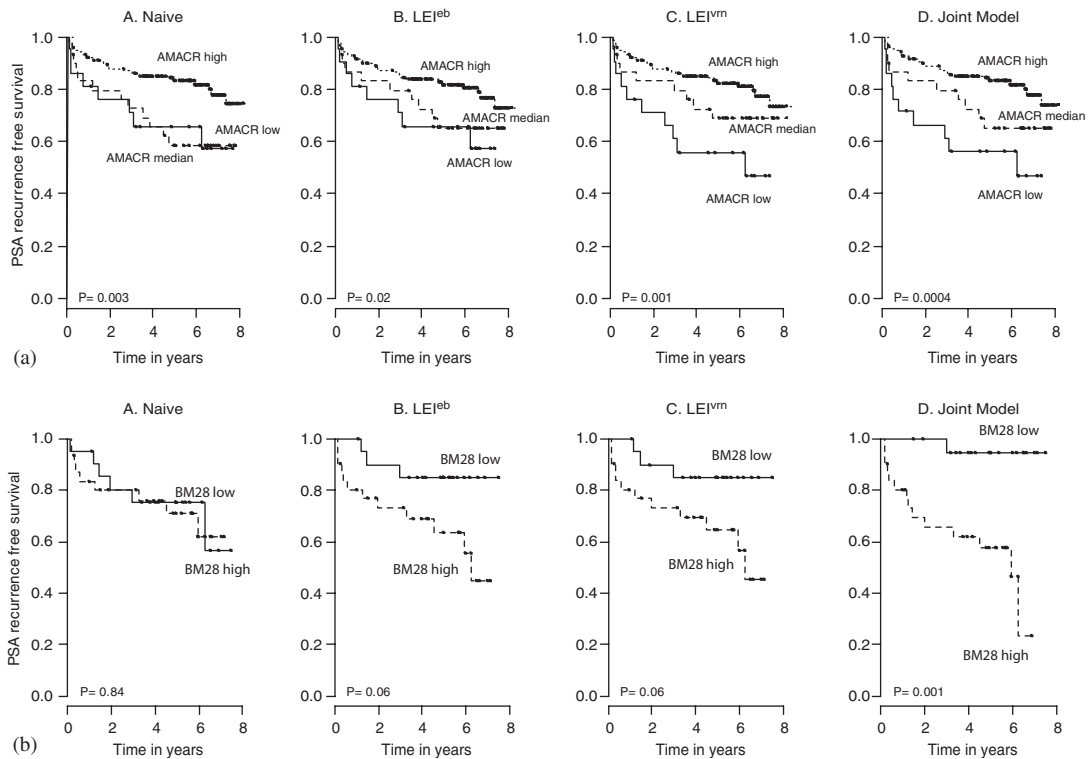


Figure 3. Kaplan–Meier plots of prostate cancer recurrence. Patients are categorized into risk groups based on the protein expression level of (a) AMACR and (b) BM28 profiled using TMAs. The expression estimates are based on the A. Naive, B. LEI^{eb}, C. LEI^{vm}, and D. Joint model.

6.4. BM28 expression and biochemical recurrence in prostate cancer

The measurement model indicates a marginal association of pathologic stage of the tumor with BM28 expression: $\hat{\theta}_{PathStage} = 0.59$ ($s.e.(\hat{\theta}) = 0.22$). The replicate number model again suggests a strong dependence on the size of the tumor: $\hat{\psi}_{Tumorsize} = 1.12$ ($s.e.(\hat{\psi}) = 0.43$).

In Table II, the differences under various models are more discernable in these data sets. First, the Weibull and piecewise exponential model overall generate slightly different results given a small sample size ($n = 52$). Second, the EB estimate differs substantially from the FB LEI estimate. It is probably due to the large uncertainties in the parameter estimates that determine LEI^{eb}. Finally, the bias introduced by the first-order approximation is prominent here. Both the coefficient β^* and the noise ratio in this data set are much larger in magnitude compared with the AMACR data example. In this case, the two-stage methods alleviate regression attenuation, only to a limited extent. The joint model should be used for parameter estimation and associated inference.

Figure 3(b) plots the Kaplan–Meier curves using different expression estimates. The 5-year PSA recurrence-free survival probability is 0.95 ($s.e. = 0.05$) versus 0.58 ($s.e. = 0.10$) for low and high BM28 expression estimated by the joint model. Adjusting for measurement error in this data set

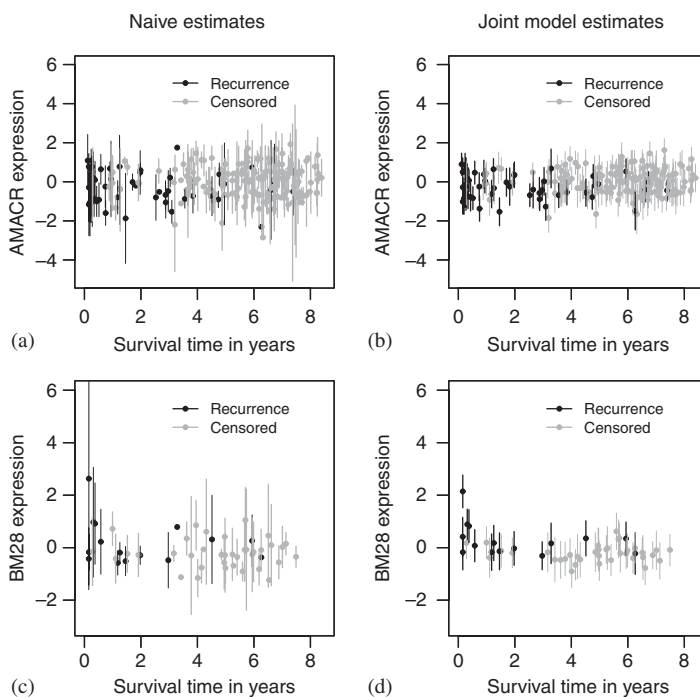


Figure 4. Comparison of the naive expression estimates ((a), (c)) and the joint model expression estimates ((b), (d)). The top panel depicts the comparison in the AMACR data, the bottom panel depicts the comparison in the BM28 data. The survival times are plotted on the x -axis.

has made a dramatic change in the conclusion about the prognostic value of BM28 compared with the naive method.

6.5. Improved expression estimates

Figure 4 compares the naive and the joint model expression estimates, plotted against the survival time on the x -axis. The mean expression \pm two standard deviations is plotted for each individual. Two improvements under the joint model are clear: (1) the noise, represented by the error bars, is greatly reduced via the joint modeling and (2) the mean expression levels are distributed more tightly around a regression line, accentuating the relationship between the TMA expression data and survival time.

7. DISCUSSION

In TMA data analysis, statistical methods often focus on downstream models in predicting disease outcome assuming \bar{X}_i is a sufficient expression summary measure [28, 29]. Relatively little attention has been paid to the modeling of within-tumor variation in these TMA experiments. As we have shown in this paper with real data examples, analysis ignoring intra-tumor variation can lead to

false negative results, causing a tremendous waste of valuable tissue resource and experimental costs. In this study, we propose and compare a two-stage approach and a joint model to analyze TMA data in a measurement error framework. The two-stage approach computes a LEI in the first stage and then associates the LEI with patient survival information using a proportional hazards model in the second stage. We further adjust for clinical/pathological covariates Z_i and the number of repeated measures (K_i, r_i) per each tumor to improve the efficiency of the expression estimates.

Although the two-stage approach is attractive for its simplicity, interpretability and straightforward computation, practitioners need to be aware of the associated statistical limitations. As we discussed earlier, the regression calibration approach is an approximation of the true model which reduces bias relative to the naive method but still leads to under-estimated β^* . In addition, the survival information is not used in the first stage to compute LEI, which can cause bias and efficiency loss in estimating β^* in the second stage. Finally, the uncertainty of estimating the LEI quantity is not assimilated in the second stage, leading to over-optimistic standard error estimates. For a review of this topic, see Tsiatis and Davidian [23]. In simulation studies, we show that the two-stage methods reduce bias relative to the naive approach in estimating the Cox regression coefficient, but still lead to under-estimated β^* . The joint analysis, however, outperforms the two-stage methods in terms of the bias and coverage property in various simulated scenarios. In case studies, we applied the error methods to prostate TMA data sets where the two-stage approach and the joint analysis generate (1) very similar results in the AMACR data set and (2) disparate results in the BM28 data set. Although the first case study demonstrates a scenario where the two-stage approach is a good approximation to the true model, the second one suggests where such approximation is not so accurate (the higher-order term cannot be ignored when β^* is a large value). Inference should be based on the joint analysis whenever results differ.

A final notion is regarding the impact of the technology advancement on TMA data analysis. Recent advances in quantitative assessment of the immunohistochemical staining provide precise, objective, and reproducible protein expression measurements. Compared with the conventional pathologist scoring on an ordinal scale, the Chromavision system used in our data examples enables quantification of the antigen level on a continuous scale, free of the subjectivity associated with pathologist-based visual scoring system. We should point out that the Chromavision system not only quantifies the amount of protein expression but also provides a more accurate measure of the percentage of tissue that express the protein. A statistical problem of interest is to extend our model to incorporate the percentage information when summarizing the expression profile of a tumor. Another quantitative system, AQUA [30], which stands for automated quantitative analysis, measures fluorescence signals, leading to higher sensitivity to very low antibody concentrations. In addition, it allows the separation of tumor from stromal elements and the sub-cellular localization of signals for a co-localization of the antigens in different cell compartments. As the technology is becoming more and more refined, statistical models underpinning both biological and experimental issues are in great need.

ACKNOWLEDGEMENTS

We thank Dr Arul Chinnaiyan at the Pathology and Urology Department at the University of Michigan Medical School and Dr Mark Rubin at the Pathology Department of Weill Medical College of Cornell

University for making the prostate TMA data sets available. We also thank the reviewers for their helpful comments to improve the manuscript.

REFERENCES

1. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature Medicine* 1998; **4**:844–847.
2. Divito KA, Berger AJ, Camp RL, Dolled-Filhart M, Rimm DL, Kluger HM. Automated quantitative analysis of tissue microarrays reveals an association between high bcl-2 expression and improved outcome in melanoma. *Cancer Research* 2004; **64**:8773–8777.
3. Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, Kurdistani SK. Global histone modification patterns predict risk of prostate cancer recurrence. *Nature* 2005; **435**:1262–1266.
4. Rubin MA, Bismar TA, Andr en O, Mucci L, Kim R, Shen R, Ghosh D, Wei J, Chinnaiyan AM, Adami H, Kantoff PJJ. Decreased α -methylacyl CoA racemase expression in localized prostate cancer is associated with an increased rate of biochemical recurrence and cancer-specific death. *Cancer Epidemiology and Biomarkers Prevention* 2005; **14**:1424–1431.
5. Li C, Wang WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proceedings of the National Academy of Sciences of the United States of America* 2001; **98**:31–36.
6. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix genechip probe level data. *Nucleic Acids Research* 2003; **31**:1–8.
7. Etzioni R, Hawley S, Billheimer D, True LD, Knudsen B. Analyzing patterns of staining in immunohistochemical studies: application to a study of prostate cancer recurrence. *Cancer Epidemiology and Biomarkers Prevention* 2005; **14**:1040–1046.
8. Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* 1995; **90**:27–37.
9. Wulfsohn MS, Tsiatis AA. A joint modeling for survival and longitudinal data measured with error. *Biometrics* 1997; **53**:330–339.
10. Faucett CJ, Thomas DC. Simultaneously modeling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* 1996; **15**:1663–1685.
11. Henderson R, Diggle P, Dobson A. Joint modeling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**:465–480.
12. Wang Y, Taylor JMG. Joint modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* 2001; **96**:895–905.
13. Xu J, Zeger SL. Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* 2001; **50**:375–387.
14. Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* 2003; **59**:221–228.
15. Guo X, Carlin B. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician* 2004; **58**:1–9.
16. Tadesse MG, Ibrahim JG, Gentleman R, Chiaretti S, Ritz J, Foa R. Bayesian error-in-variable survival model for the analysis of genechip arrays. *Biometrics* 2005; **61**:488–497.
17. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 1982; **69**:331–342.
18. Carroll RJ, Ruppert D, Stefanski LA. *Measurement Error in Nonlinear Models*. Chapman & Hall: London, 1995.
19. Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 1977; **72**:320–338.
20. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.
21. Lindstrom MJ, Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics* 1990; **46**:673–687.
22. McCulloch CE. Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association* 1994; **89**:330–335.
23. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 2004; **14**:809–834.

24. Spiegelhalter D, Thomas A, Best N, Lunn D. *WinBUGS 1.4 Manual*, 2003. Available from: <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>.
25. Thomas A. *BRugs: An R interface to OpenBUGS. Version 1.0 User's Manual*, 2004. Available from: <http://mathstat.helsinki.fi/openbugs/>.
26. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Research* 2002; **62**:4427C33.
27. Bismar T, Demichelis F, Riva A, Kim R, Varambally S, He L. Defining aggressive prostate cancer using a 12-gene model. *Neoplasia* 2006; **8**:59–68.
28. Horvath S, Liu X, Huang Y, Kim HL, Seligson DB. Forest-based methods for analyzing tissue microarray. *ASA Proceedings of the Joint Statistical Meetings, 1863–1869*. American Statistical Association: Alexandria, VA, 2003.
29. Liu X, Minin V, Huang Y, Seligson DB, Horvath S. Statistical methods for analyzing tissue microarray data. *Journal of Biopharmaceutical Statistics* 2004; **14**:671–685.
30. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nature Medicine* 2002; **8**:1323–1328.