Biometrika (2015), **102**, 4, *pp.* 907–923
Printed in Great Britain
doi: 10.1093/biomet/asv047
Advance Access publication 4 November 2015

Diagnostic measures for the Cox regression model with missing covariates

By HONGTU ZHU, JOSEPH G. IBRAHIM

Department of Biostatistics, University of North Carolina at Chapel Hill, 3109 McGavran-Greenberg Hall, CB #7420, Chapel Hill, North Carolina 27516, U.S.A.
hzhu@bios.unc.edu ibrahim@bios.unc.edu

and MING-HUI CHEN

Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, Connecticut 06269, U.S.A.
ming-hui.chen@uconn.edu

Summary

We investigate diagnostic measures for assessing the influence of observations and model misspecification on the Cox regression model when there are missing covariate data. Our diagnostics include case-deletion measures, conditional martingale residuals, and score residuals. The Q-distance is introduced to examine the effects of deleting individual observations on the estimates of finite- and infinite-dimensional parameters. Conditional martingale residuals are used to construct goodness-of-fit statistics for testing misspecification of the model assumptions. A resampling method is developed to approximate the p-values of the goodness-of-fit statistics. We conduct simulation studies to evaluate our methods, and analyse a real dataset to illustrate their use.

Some key words: Case-deletion measure; Conditional martingale residual; Goodness-of-fit statistic; Model misspecification.

1. Introduction

In surveys, clinical trials and longitudinal studies, complete data are often not available for every subject. There is a very large literature on statistical methods for missing data. These methods, however, depend strongly on the missing-data mechanism and on other distributional and modelling assumptions, and can be very sensitive to them. For this reason, analyses are carried out to check the sensitivity of the parameter estimates to assumptions. See, for example, Verbeke et al. (2001), Jansen et al. (2003), Troxel et al. (2004), Copas & Eguchi (2005) and Daniels & Hogan (2008).

Diagnostic measures such as martingale residuals and Cook's distance have been widely used to identify influential observations and to test for model misspecification in survival models (Storer & Crowley, 1985; Pettitt & Daud, 1989; Therneau et al., 1990; Escobar & Meeker, 1992; Henderson & Oman, 1993; Lin et al., 1993; Barlow, 1997; Marzec & Marzec, 1997; Klein & Moeschberger, 2003; Martinussen & Scheike, 2006). For instance, Pettitt & Daud (1989) applied the local influence method of Cook (1986) to the proportional hazards model and derived several useful diagnostics. Martingale residuals have been widely used to construct goodness-of-fit

statistics to examine the functional form of a covariate and the proportional hazards assumption (Barlow & Prentice, 1988; Therneau et al., 1990; Lin et al., 1993). However, to the best of our knowledge, almost no work exists on developing diagnostic measures in the Cox regression model (Cox, 1972, 1975) with missing covariate data, except for Scheike et al. (2010).

2. COX REGRESSION WITH MISSING COVARIATES

2·1. *Model set-up*

Consider n observations $(x_1, z_1, r_1, y_1, \delta_1), \ldots, (x_n, z_n, r_n, y_n, \delta_n)$ which are independent realizations of (X, Z, R, Y, Δ), where $Y = T \wedge C$ is the minimum of the censoring time C and the survival time T, $\Delta = 1(T \leqslant Y)$, which equals 1 if the observed event is a failure and 0 otherwise, and each X is a $p_1 \times 1$ vector of completely observed covariates; each $Z = (Z_{\mathrm{m}}, Z_{\mathrm{o}})$ is a $p_2 \times 1$ vector of partially observed covariates, where Z_{m} and Z_{o} denote the missing and observed components of Z, respectively. Here R is a $p_2 \times 1$ random vector whose kth component, R_k, equals 1 if Z_k is observed and 0 if Z_k is missing, where Z_k denotes the kth component of Z. Under a general missing-data mechanism, it is common to specify the joint density of (X, Z, R, Y, Δ) as a product of three conditional densities as follows:

$$p(X, Z, R, Y, \Delta) = p(Y, \Delta \mid X, Z)\, p(X, Z)\, p(R \mid X, Z, Y, \Delta). \tag{1}$$

The conditional density of $(Y, \Delta) = (y_i, \delta_i)$ given $v_i = (x_i, z_i)$ is assumed to be

$$p(y_i, \delta_i \mid v_i) \propto \lambda_t(y_i \mid v_i)^{\delta_i} S_t(y_i \mid v_i) \lambda_c(y_i \mid v_i)^{1-\delta_i} S_c(y_i \mid v_i) \quad (i = 1, \ldots, n), \tag{2}$$

where $\lambda_t(\cdot)$ and $S_t(\cdot)$ are the hazard and survivor functions of the failure time and $\lambda_c(\cdot)$ and $S_c(\cdot)$ are the hazard and survivor functions of the censoring time. We also assume the Cox model for the failure time,

$$\lambda_t(y_i \mid v_i) = h_0(y_i) \exp(v_i^{\mathrm{T}} \beta), \quad S_t(y_i \mid v_i) = \exp\{-\exp(v_i^{\mathrm{T}} \beta) H_0(y_i)\}, \tag{3}$$

where $h_0(y)$ is a baseline hazard function and $H_0(y) = \int_0^y h_0(u)\, \mathrm{d}u$.

We need to specify a joint distribution for the covariate vector $V = (X, Z)$. It is assumed that $p(v_i; \alpha) \propto p(z_i \mid x_i; \alpha)\, p(x_i)$, where α contains all the unknown parameters in $p(v_i; \alpha)$. Since the x_i are fully observed, it is not necessary to specify a distribution for X. We follow Lipsitz & Ibrahim (1996) to model $p(z_i \mid x_i; \alpha)$ as the product of one-dimensional conditional distributions. We need to consider different ways of modelling the missing-data mechanism $p(r_i \mid v_i, y_i, \delta_i; \xi)$ (Ibrahim et al., 1999), where ξ contains all the unknown parameters. It is common to use logistic regression models for the binary variables in r_i.

We calculate the conditional distribution of $Z_{\mathrm{m}} = z_{\mathrm{m},i}$ given $D_{\mathrm{o}} = d_{\mathrm{o},i}$ as

$$p(z_{\mathrm{m},i} \mid d_{\mathrm{o},i}) = \frac{\lambda_t(y_i \mid v_i)^{\delta_i} S_t(y_i \mid v_i) \lambda_c(y_i \mid v_i)^{1-\delta_i} S_c(y_i \mid v_i) p(z_i \mid x_i; \alpha)\, p(r_i \mid v_i, y_i, \delta_i; \xi)}{\int \lambda_t(y_i \mid v_i)^{\delta_i} S_t(y_i \mid v_i) \lambda_c(y_i \mid v_i)^{1-\delta_i} S_c(y_i \mid v_i) p(z_i \mid x_i; \alpha)\, p(r_i \mid v_i, y_i, \delta_i; \xi)\, \mathrm{d}z_{\mathrm{m},i}},$$

where $d_{\mathrm{o},i} = (x_i, z_{\mathrm{o},i}, r_i, y_i, \delta_i)$. If the censoring time does not depend on the missing data and all the unknown parameters, then we can drop the hazard and survivor functions of the censoring

times from the model. Moreover, if the missing data are missing at random, then

$$p(z_{\mathrm{m},i} \mid d_{\mathrm{o},i}) = \frac{\lambda_t(y_i \mid v_i)^{\delta_i} S_t(y_i \mid v_i) p(z_i \mid x_i; \alpha)}{\int \lambda_t(y_i \mid v_i)^{\delta_i} S_t(y_i \mid v_i)(z_i \mid x_i; \alpha)\, \mathrm{d}z_{\mathrm{m},i}}.$$

The expectation-maximization algorithm is a popular technique for obtaining the maximum likelihood estimates of $\eta = \{h_0(\cdot), \gamma\}$, denoted by $\hat{\eta} = \{\hat{h}_0(\cdot), \hat{\gamma}\}$, in the Cox regression model with missing covariate data (Chen & Little, 1999; Herring & Ibrahim, 2001), where $\gamma = (\beta^{\mathrm{T}}, \alpha^{\mathrm{T}}, \xi^{\mathrm{T}})^{\mathrm{T}}$. Let D_{c} and D_{o} denote the complete and observed data, respectively. We calculate the nonparametric maximum likelihood estimator of $H_0(\cdot)$, which is a step function with jumps only at the y_i such that $\delta_i = 1$ $(i = 1, \ldots, n)$. Without loss of generality, we assume that $y_1, \ldots, y_d$ are d distinct failure times. At the sth step of the expectation-maximization algorithm, given $\eta^{(s)}$, the expectation step involves evaluating the Q-function $Q(\eta \mid \eta^{(s)}) = E\{L_{\mathrm{c}}(\eta \mid D_{\mathrm{c}}) \mid D_{\mathrm{o}}, \eta^{(s)}\}$, which has the form

$$\begin{aligned} Q(\eta \mid \eta^{(s)}) &= \sum_{i=1}^{n} \int \log\big[p\{y_i, \delta_i \mid x_i, z_i; \beta, h_0(\cdot)\}\big]\, p(z_{\mathrm{m},i} \mid x_i, z_{\mathrm{o},i}, r_i, y_i, \delta_i; \eta^{(s)})\, \mathrm{d}z_{\mathrm{m},i} \\ &\quad + \sum_{i=1}^{n} \int \log\{p(x_i, z_i; \alpha)\}\, p(z_{\mathrm{m},i} \mid x_i, z_{\mathrm{o},i}, r_i, y_i, \delta_i; \eta^{(s)})\, \mathrm{d}z_{\mathrm{m},i} \\ &\quad + \sum_{i=1}^{n} \int \log\{p(r_i \mid x_i, z_i, y_i, \delta_i; \xi)\}\, p(z_{\mathrm{m},i} \mid x_i, z_{\mathrm{o},i}, r_i, y_i, \delta_i; \eta^{(s)})\, \mathrm{d}z_{\mathrm{m},i} \\ &= Q_1\{\beta, h_0(\cdot) \mid \eta^{(s)}\} + Q_2(\alpha \mid \eta^{(s)}) + Q_3(\xi \mid \eta^{(s)}), \end{aligned} \tag{4}$$

where $L_{\mathrm{c}}(\eta \mid D_{\mathrm{c}}) = \log p(D_{\mathrm{c}}; \eta)$ is the complete-data loglikelihood function. The maximization step consists of maximizing $Q_1\{\beta, h_0(\cdot) \mid \eta^{(s)}\}$, $Q_2(\alpha \mid \eta^{(s)})$ and $Q_3(\xi \mid \eta^{(s)})$ separately (Chen & Little, 1999; Herring & Ibrahim, 2001).

Our main interest is in making valid inferences about β and $H_0(y)$, and this requires the correct specification of all three levels of the assumptions in (1); otherwise there may be serious bias in estimating β and $H_0(\cdot)$. Therefore, it is crucial to assess the potential misspecification of all the assumptions in (1).

2·2. *Assumptions*

The following assumptions are needed to facilitate the development of our methods, although they may not be the weakest possible conditions.

Assumption 1. The C_i and T_i given $V = v_i$ are independent, and the hazard and survivor functions of C_i do not depend on $z_{\mathrm{m},i}$ and η.

Assumption 2. The true value $(\alpha_*, \beta_*, \xi_*)$ of (α, β, ξ) is an interior point of the compact parameter space of (α, β, ξ).

Assumption 3. The functions $\log p(v; \alpha)$ and $\log p(r \mid x, z, y, \delta; \xi)$ are twice continuously differentiable in γ, and the absolute values of their first- and second-order derivatives are dominated by a function $B(d)$. For each i, $B(d_i)$ is integrable such that $\sup_\eta E\{B(d_i)^2 \mid D_{\mathrm{o}}; \eta\} = O_{\mathrm{p}}(1)$. Moreover, v is bounded, $p(v; \alpha)$ is uniformly bounded and identifiable, and $\mathrm{var}(v)$ and $\int\{-\partial_\alpha^2 \log p(v; \alpha_*)\} p(v; \alpha_*)\, \mathrm{d}v$ are positive definite.

Assumption 4. Let τ be a finite time-point at which any individual still under study is censored. Assume that $\mathrm{pr}(Y \geqslant \tau) > 0$. The function $H_0(t) = \int_0^t h_0(s)\,\mathrm{d}s$ is an absolutely continuous nondecreasing function such that $H_0(0) = 0$ and $H_0(\tau) < \infty$. Moreover, $h_0(s) \geqslant 0$ is twice continuously differentiable.

Assumption 5. The missing covariate data are missing at random, i.e., $\mathrm{pr}(r \mid x, z, y, \delta) = \mathrm{pr}(r \mid x, z_{\mathrm{o}}, y, \delta)$. In addition, the fully observed complete covariates can be observed for all possible covariate values; that is, $\mathrm{pr}(r = 1_{p_2} \mid x, z, y, \delta) > 0$ holds for almost all (x, z) and almost all $y \in [t_1, t_2]$ such that $H_0(t_1) \neq H_0(t_2)$, where 1_{p_2} is a $p_2 \times 1$ vector of ones.

Assumption 6. The probability function $F_\varphi(\mathrm{d}t)\,\mathrm{d}\varphi$ is absolutely continuous with respect to the Lebesgue measure on $\Pi = \{\varphi \in \mathbb{R}^{p_1} : \varphi^{\mathrm{T}}\varphi = 1\} \times [-\infty, \infty]$.

Assumption 7. As $n \to \infty$, for any sequences $\{(\varphi_n, u_n, t_n)\}$ and $\{(\varphi_{n,1}, u_{n,1}, t_{n,1})\}$, $\rho_n(\varphi_n, u_n, t_n; \varphi_{n,1}, u_{n,1}, t_{n,1})$ converges to zero when $\rho(\varphi_n, u_n, t_n; \varphi_{n,1}, u_{n,1}, t_{n,1}) \to 0$. Moreover, $\rho(\varphi, u, t; \varphi_1, u_1, t_1)$ is the limit of $\rho_n(\varphi_n, u_n, t_n; \varphi_{n,1}, u_{n,1}, t_{n,1})$, which is defined as

$$\left(n^{-1} \sum_{i=1}^{n} E\left[\left\{ R_i(t_n) 1(\varphi_n^{\mathrm{T}} x_i \leqslant u_n) - R_i(t_{n,1}) 1(\varphi_{n,1}^{\mathrm{T}} x_i \leqslant u_{n,1}) \right\}^2 \right] \right)^{1/2},$$

where $R_i(t)$ is a conditional martingale residual to be introduced later.

Assumption 8. For any small $a_0 > 0$,

$$\sup_{(\alpha,\varphi,t) \in \mathcal{A} \times \Pi} \mathrm{pr}\left[-\delta < \{v_i(\alpha)^{\mathrm{T}}\varphi - t\} / V_i(x_i, z_{\mathrm{o},i}) < \delta \right] \leqslant C_0\, \delta^{c_1},$$

where C_0 and c_1 are two positive scalars, $\mathcal{A} = \{\alpha : \|\alpha - \alpha_*\| \leqslant a_0\}$, and $V_i(x_i, z_{\mathrm{o},i})^2 = \sup_{\alpha \in \mathcal{A}} \|\partial_\alpha v_i(\alpha)\|^2 + \sup_{\alpha \in \mathcal{A}} \|v_i(\alpha)\|^2 + 1$. Moreover,

$$\begin{aligned} &v_i(\hat{\alpha}) \\ &= \left\{ x_i, r_{i1} z_{i1} + (1 - r_{i1}) E(z_{i1} \mid x_i, z_{\mathrm{o},i}; \hat{\alpha}), \ldots, r_{ip_2} z_{ip_2} + (1 - r_{ip_2}) E(z_{ip_2} \mid x_i, z_{\mathrm{o},i}; \hat{\alpha}) \right\}^{\mathrm{T}}. \end{aligned}$$

Assumption 9. Let $\lambda_{\min}(\cdot)$ be the smallest eigenvalue of a matrix. For a fixed $\epsilon_0 > 0$,

$$\begin{gathered} n^{-1} \left[Q\{\hat{\gamma}, \hat{h}_0(\cdot) \mid \hat{\eta}\} - \sup_{\|\gamma - \hat{\gamma}\| = \epsilon_0} Q\{\gamma, \hat{h}_0(\cdot) \mid \hat{\eta}\} \right] = C_0 + o_{\mathrm{p}}(1), \\ \sup_{\|\gamma - \hat{\gamma}\| \leqslant \epsilon_0} \left\| n^{-1} \partial_\gamma^2 Q\{\gamma, \hat{h}_0(\cdot) \mid \hat{\eta}\} - A(\gamma) \right\| = o_{\mathrm{p}}(1), \end{gathered}$$

where $\min_{\|\gamma - \hat{\gamma}\| \leqslant \epsilon_0} \lambda_{\min}\{A(\gamma)^2\} > 0$ and C_0 is a positive scalar.

Assumptions 1–5 have been used to establish consistency and asymptotic normality of the nonparametric maximum likelihood estimator in a proportional hazards regression model with covariates missing at random (Chen & Little, 1999). Assumption 6 is required to establish the asymptotic distributions of the Cramer–von Mises test statistics introduced below. Assumption 7 is needed in order to invoke the central limit theory for sums of independent but not identically distributed stochastic processes (Pollard, 1990; van der Vaart & Wellner, 1996; Kosorok, 2007). Assumption 8 is required to invoke Ossiander's entropy conditions (Ossiander, 1987;

Andrews, 1994). Assumption 9 is needed to establish the asymptotic accuracy of approximating case-deletion measures introduced below.

3. Diagnostic measures

3·1. *Case-deletion influence measures*

To quantify the effects of deleting the ith observation on the maximum likelihood estimate $\hat{\eta}$ of η, it is common to compute the maximum likelihood estimate of η for a subsample $D_{c[i]}$, obtained upon deleting the ith observation $d_i = (y_i, \delta_i, v_i, r_i)$ from $D_c = \{D_o, D_m\} = \{(y_j, \delta_j, c_j, r_j) : j = 1, \ldots, n\}$, where D_o and D_m denote the observed and missing data, respectively. However, it is computationally intensive to directly maximize the likelihood function based on the subsample $D_{c[i]}$ for each i. Instead, we define $Q_{[i]}(\eta \mid \hat{\eta})$ as $Q_{[i]}(\eta \mid \hat{\eta}) = E\{L_c(\eta \mid D_{c[i]}) \mid D_o; \hat{\eta}\}$, where $L_c(\eta \mid D_{c[i]})$ denotes the complete-data loglikelihood function for $D_{c[i]}$ and the expectation is taken with respect to $p(D_m \mid D_o; \hat{\eta})$. Similar to (4), $Q_{[i]}(\eta \mid \hat{\eta})$ equals the sum of $Q_{1[i]}\{\beta, h_0(\cdot) \mid \hat{\eta}\} = \sum_{j \neq i} E(\log[p\{y_j, \delta_j \mid x_j, z_j; \beta, h_0(\cdot)\}] \mid D_o; \hat{\eta})$, $Q_{2[i]}(\alpha \mid \hat{\eta}) = \sum_{j \neq i} E[\log\{p(x_j, z_j; \alpha)\} \mid D_o; \hat{\eta}]$ and $Q_{3[i]}(\xi \mid \hat{\eta}) = \sum_{j \neq i} E[\log\{p(r_j \mid x_j, z_j, y_j, \delta_j; \xi)\} \mid D_o; \hat{\eta}]$.

Let $\omega = (\omega_1, \ldots, \omega_n)^T$ with $\omega_k \geqslant 0$ for all k. We define $Q_1\{\omega, \beta, h_0(\cdot) \mid \hat{\eta}\}$ to be

$$\sum_{k=1}^{n} \omega_k \delta_k \{\log h_0(y_k) + E(c_k^T \beta \mid D_o; \hat{\eta})\} - \sum_{k=1}^{n} \omega_k H_0(y_k) E\{\exp(c_k^T \beta) \mid D_o; \hat{\eta}\}. \tag{5}$$

First, by substituting $\hat{h}_0(\cdot)$ into (5), we can obtain $Q_1\{\omega, \beta, \hat{h}_0(\cdot)\}$ as

$$\sum_{k=1}^{n} \delta_k \omega_k E(c_k^T \beta \mid D_o; \hat{\eta}) - \sum_{k=1}^{n} \omega_k \hat{H}_0(y_k) E\{\exp(c_k^T \beta) \mid D_o; \hat{\eta}\}.$$

We calculate $\hat{\beta}(\omega) = \arg\max_\beta Q_1\{\omega, \beta, \hat{h}_0(\cdot)\}$ and then maximize $Q_1\{\omega, \hat{\beta}(\omega), h_0(\cdot)\}$ with respect to $h_0(\cdot)$, leading to

$$\hat{h}_0(y_k \mid \beta, \omega) = \frac{\delta_k \omega_k}{\sum_{j \in R_k} \omega_j E\{\exp(c_j^T \beta) \mid D_o; \hat{\eta}\}},$$

where $R_k = \{j : y_j \geqslant y_k\}$. If $\omega = 1_n$ is an $n \times 1$ vector of ones, then $\hat{\beta}(1_n) = \hat{\beta}$ and $\hat{h}_0(y_k) = \hat{h}_0(y_k \mid \hat{\beta}, 1_n)$. Furthermore, if $\omega = 1_n - e_i$, then we define $\hat{\beta}_{[i]} = \hat{\beta}(1_n - e_i)$ and $\hat{h}_{0[i]}(y_k) = \hat{h}_0(y_k \mid \hat{\beta}_{[i]}, 1_n - e_i)$. Similarly, we define $\hat{\alpha}_{[i]}$ and $\hat{\xi}_{[i]}$ as the maximizers of $Q_{2[i]}(\alpha \mid \hat{\eta})$ and $Q_{3[i]}(\xi \mid \hat{\eta})$, respectively. Now we can calculate a one-step approximation $\hat{\eta}^1_{[i]} = \{\hat{h}^1_{0[i]}(\cdot), \hat{\beta}^1_{[i]}, \hat{\alpha}^1_{[i]}, \hat{\xi}^1_{[i]}\}$ of $\hat{\eta}_{[i]} = \{\hat{h}_{0[i]}(\cdot), \hat{\beta}_{[i]}, \hat{\alpha}_{[i]}, \hat{\xi}_{[i]}\}$ as below. We obtain the following theorem, whose proof is given in the Supplementary Material.

Theorem 1. *Under Assumptions 3 and 9,*

$$\begin{aligned}
\hat{\beta}^1_{[i]} &= \hat{\beta} - [-\partial^2_\beta Q_1\{1_n, \hat{\beta}, \hat{h}_0(\cdot)\}]^{-1} \partial^2_{\beta\omega_i} Q_1\{1_n, \hat{\beta}, \hat{h}_0(\cdot)\} = \hat{\beta}_{[i]} + o_p(n^{-1}),\\
\hat{\alpha}^1_{[i]} &= \hat{\alpha} - \{-\partial^2_\alpha Q_2(\hat{\alpha} \mid \hat{\eta})\}^{-1} E\{\partial_\alpha \log p(v_i; \hat{\alpha}) \mid D_o; \hat{\eta}\} = \hat{\alpha}_{[i]} + o_p(n^{-1}),\\
\hat{\xi}^1_{[i]} &= \hat{\xi} - \{-\partial^2_\xi Q_3(\hat{\xi} \mid \hat{\eta})\}^{-1} E\{\partial_\xi \log p(r_i \mid d_{o,i}; \hat{\xi}) \mid D_o; \hat{\eta}\} = \hat{\xi}_{[i]} + o_p(n^{-1}),\\
\hat{h}^1_{0[i]}(y_k) &= \hat{h}_0(y_k \mid \hat{\beta}^1_{[i]}, 1_n - e_i) = \hat{h}_{0[i]}(y_k) + o_p(n^{-1}).
\end{aligned} \tag{6}$$

Theorem 1 gives the one-step approximation $\hat{\eta}^1_{[i]}$ of $\hat{\eta}_{[i]}$ for each major component of η. It is straightforward to compute $\hat{\eta}^1_{[i]}$ using (6).

We introduce a Q-distance for the finite-dimensional parameter γ in the presence of an infinite-dimensional parameter $h_0(\cdot)$ to quantify the distance between the maximum likelihood estimates of γ with and without the ith observation having been deleted from the full sample (Cook & Weisberg, 1982; Zhu et al., 2001). The Q-distance for the ith subject is defined as

$$\mathrm{QD}_i(M) = (\hat{\gamma}^1_{[i]} - \hat{\gamma})^{\mathrm{T}} M (\hat{\gamma}^1_{[i]} - \hat{\gamma}),$$

where M is a positive-definite matrix. According to (4), we assume that $-M = \mathrm{diag}[\partial^2_{\beta} Q_1\{1_n, \hat{\beta}, \hat{h}_0(\cdot)\}, \partial^2_{\alpha} Q_2(\hat{\alpha} \mid \hat{\eta}), \partial^2_{\xi} Q_3(\hat{\xi} \mid \hat{\eta})]$. Thus, QD_i can be decomposed into a sum of three diagnostic measures based on (1)–(3); that is, $\mathrm{QD}_i = \mathrm{QD}_{i,1} + \mathrm{QD}_{i,2} + \mathrm{QD}_{i,3}$ where

$$\begin{aligned}
\mathrm{QD}_{i,1} &= [\partial^2_{\omega_i \beta} Q_1\{1_n, \hat{\beta}, \hat{h}_0(\cdot)\}]^{\mathrm{T}} [-\partial^2_{\beta} Q_1\{1_n, \hat{\beta}, \hat{h}_0(\cdot)\}]^{-1} \partial^2_{\beta \omega_i} Q_1\{1_n, \hat{\beta}, \hat{h}_0(\cdot)\},\\
\mathrm{QD}_{i,2} &= E\{\partial_{\alpha} \log p(v_i; \hat{\alpha}) \mid D_{\mathrm{o}}; \hat{\eta}\}^{\mathrm{T}} \{-\partial^2_{\alpha} Q_2(\hat{\alpha} \mid \hat{\eta})\}^{-1} E\{\partial_{\alpha} \log p(v_i; \hat{\alpha}) \mid D_{\mathrm{o}}; \hat{\eta}\},\\
\mathrm{QD}_{i,3} &= E\{\partial_{\xi} \log p(r_i \mid d_{\mathrm{o},i}; \hat{\xi}) \mid D_{\mathrm{o}}; \hat{\eta}\}^{\mathrm{T}} \{-\partial^2_{\xi} Q_3(\hat{\xi} \mid \hat{\eta})\}^{-1} E\{\partial_{\xi} \log p(r_i \mid d_{\mathrm{o},i}; \hat{\xi}) \mid D_{\mathrm{o}}; \hat{\eta}\}.
\end{aligned}$$

Intuitively, $\mathrm{QD}_{i,1}$, $\mathrm{QD}_{i,2}$ and $\mathrm{QD}_{i,3}$ are associated with the effects of removing the ith observation on the assumptions of $p\{y_i, \delta_i \mid c_i; \beta, h_0(\cdot)\}$, $p(v_i; \alpha)$ and $p(r_i \mid v_i, y_i, \delta_i; \xi)$. If QD_i is large, then the ith observation is influential. Similarly, we can quantify the effects of deleting two or more observations on $\hat{\eta}$ (Cook & Weisberg, 1982), but for simplicity we omit those details here.

We also define a distance function of $\hat{h}_0(\cdot)$ and $\hat{h}^1_{0[i]}(\cdot)$ to quantify the effect of deleting the ith observation on the infinite-dimensional parameter $h_0(\cdot)$. Let $\|\cdot\|_{\infty}$ denote the sup-norm for functions. Specifically, we define

$$\mathrm{QD}_{i,h_0(\cdot)} = \max_{1 \leqslant j \leqslant n} \left| \sum_{k=1}^{n} Y_k(y_j) \{\hat{h}_0(y_k) - \hat{h}^1_{0[i]}(y_k)\} \right| = \left\| \hat{H}_0 - \hat{H}^1_{0[i]} \right\|_{\infty},$$

where $Y_k(u) = 1(y_k \geqslant u)$, $\hat{H}_0(y) = \sum_{y_j \leqslant y} \hat{h}_0(y_j)$ and $\hat{H}^1_{0[i]}(y) = \sum_{y_j \leqslant y} \hat{h}^1_{0[i]}(y_j)$.

A challenging problem is the quantification of the magnitude of these case-deletion measures for detecting influential observations. A common approach is to sort these measures for all observations and then classify observations with larger measures as influential. However, this method may not identify truly influential observations, and it does not reveal why an observation is influential. To address this issue, we introduce a detection probability of being influential for each observation and for any case-deletion measure. The key idea is to measure the standardized influential level of each observation for a case-deletion measure under the assumption that (1) is the true data generator. We compute the detection probabilities of all observations based on the fitted model $p(d_i; \hat{\eta})$ as follows. First, we use a semi-bootstrap method, described in the Supplementary Material, to generate multiple bootstrapped datasets. Then, for each bootstrapped dataset, we calculate all of the case-deletion diagnostic measures across all observations. For each observation, the detection probability is calculated as the proportion of the bootstrapped case-deletion diagnostic measures that are smaller than the corresponding observed case-deletion diagnostic measure. Observations with large detection probabilities, say 0·95 or greater, can be regarded as influential.

3·2. *Residuals*

We consider two types of residuals: conditional martingale residuals and score residuals for the Cox regression model with missing covariates. When there are no missing covariates, the martingale residual for the ith observation at time t is defined as

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) \exp(v_i^{\mathrm{T}}\beta) h_0(u)\, \mathrm{d}u,$$

where $N_i(t) = \delta_i 1(T_i \leqslant t)$. However, since $z_{\mathrm{m},i}$ is missing, $M_i(t)$ cannot be directly calculated for cases with missing covariates. Although there are many ways of integrating out $z_{\mathrm{m},i}$, we define a conditional martingale residual for the ith observation at t by

$$R_i(t) = N_i(t) - \int_0^t Y_i(u) E\{\exp(v_i^{\mathrm{T}}\beta) \mid d_{\mathrm{o},i}\} h_0(u)\, \mathrm{d}u \quad (i = 1, \ldots, n), \tag{7}$$

where $d_{\mathrm{o},i} = (y_i, \delta_i, x_i, z_{\mathrm{o},i}, r_i)$ and the expectation is taken with respect to $p(z_{\mathrm{m},i} \mid d_{\mathrm{o},i}; \eta)$. If there are no missing covariates in z_i, then $R_i(t)$ reduces to $M_i(t)$. Thus, $R_i(t)$ can be regarded as a generalization of the martingale residuals used in Cox regression. Computationally, the conditional expectation in (7) can easily be calculated using Markov chain Monte Carlo methods (Chen et al., 2000). Then $R_i(t)$ evaluated at $\hat{\eta}$ is given by

$$\hat{R}_i(t) = N_i(t) - \int_0^t Y_i(u) E\{\exp(v_i^{\mathrm{T}}\hat{\beta}) \mid d_{\mathrm{o},i}; \hat{\eta}\} \hat{h}_0(u)\, \mathrm{d}u.$$

In particular, when $t = \tau = \sup\{u : \mathrm{pr}\{Y(u) = 1\} > 0\}$, i.e., the end time of the study, we can obtain the corresponding conditional martingale residual as follows:

$$\hat{R}_i = \hat{R}_i(\tau) = \delta_i - \hat{r}_i = \delta_i - \int_0^{y_i} E\{\exp(v_i^{\mathrm{T}}\hat{\beta}) \mid d_{\mathrm{o},i}; \hat{\eta}\} \hat{h}_0(u)\, \mathrm{d}u,$$

where $\hat{r}_i$ is a generalization of the Cox–Snell residual in the case of missing covariates (Cox & Snell, 1968).

Turning to the score residual, we define $S^{(r)}(\beta, u; \hat{\eta}) = n^{-1} \sum_{i=1}^n Y_i(u) E\{\exp(v_i^{\mathrm{T}}\beta) v_i^{\otimes r} \mid D_{\mathrm{o}}; \hat{\eta}\}$ for $r = 0, 1, 2$, where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = a\, a^{\mathrm{T}}$ for a vector a. The score function associated with β is

$$\begin{aligned} \partial_\beta Q(\hat{\eta} \mid \hat{\eta}) &= \sum_{i=1}^n \left[\delta_i E(v_i \mid d_{\mathrm{o},i}; \hat{\eta}) - \hat{H}_0(y_i) E\{v_i \exp(v_i^{\mathrm{T}}\hat{\beta}) \mid d_{\mathrm{o},i}; \hat{\eta}\}\right] \\ &= \sum_{i=1}^n \int_0^\infty U_i(u, \hat{\eta})\, \mathrm{d}N_i(u), \end{aligned}$$

where $U_i(u; \eta) = \{U_{i,1}(u; \eta)^{\mathrm{T}}, U_{i,2}(u; \eta)^{\mathrm{T}}\}^{\mathrm{T}} = E(v_i \mid d_{\mathrm{o},i}; \eta) - S^{(1)}(\beta, u; \eta)/S^{(0)}(\beta, u; \eta)$, with $U_{i,1}(u; \eta)$ denoting the first p_1 components of $U_i(u; \eta)$ associated with x_i. Further, we can define a score process for β,

$$U(t \mid \eta) = \{U_1(t \mid \eta)^{\mathrm{T}}, U_2(t \mid \eta)^{\mathrm{T}}\}^{\mathrm{T}} = \sum_{i=1}^n \int_0^t U_i(u; \eta)\, \mathrm{d}N_i(u),$$

where $U_1(t \mid \eta)$ denotes the first p_1 components of $U(t \mid \eta)$ associated with x_i. Finally, we have $0 = \partial_\beta Q\{\hat{\beta}, \hat{h}_0(\cdot) \mid \hat{\eta}\} = U(\tau \mid \hat{\eta}) = \sum_{i=1}^n \hat{S}_i$, where $\hat{S}_i = (\hat{s}_{i1}, \ldots, \hat{s}_{ip})$ is given by

$$\begin{pmatrix} \hat{S}_{i,1} \\ \hat{S}_{i,2} \end{pmatrix} = \delta_i \begin{pmatrix} x_i \\ E(z_i \mid d_{\mathrm{o},i}; \hat{\eta}) \end{pmatrix} - \hat{H}_0(y_i) \exp(x_i^{\mathrm{T}} \hat{\beta}_1) \begin{pmatrix} x_i E\{\exp(z_i^{\mathrm{T}} \hat{\beta}_2) \mid d_{\mathrm{o},i}; \hat{\eta}\} \\ E\{z_i \exp(z_i^{\mathrm{T}} \hat{\beta}_2) \mid d_{\mathrm{o},i}; \hat{\eta}\} \end{pmatrix},$$

with $\hat{S}_{i,1}$ being the first $p_1 \times 1$ subvector of $\hat{S}_i$ associated with β_1. Score residuals are useful tools in detecting influential observations and in assessing model assumptions (Therneau et al., 1990). As with the case-deletion diagnostic measures, we can use the semi-bootstrap method to generate random samples and then calculate the detection probabilities of $|\hat{s}_{ik}|$ for $k = 1, \ldots, p$.

We study several properties of the proposed conditional martingale residuals and score residuals. Through a better understanding of the properties of these residuals, we can develop both formal and informal diagnostic tools to examine the adequacy of the Cox regression model with missing covariates.

THEOREM 2. *Suppose that Assumption 3 holds. Then:*

(i) $E\{R_i(t) \mid x_i, z_{\mathrm{o},i}\} = E\{R_i(t) \mid x_i\} = E\{R_i(t)\} = 0$;
(ii) *in general,* $E\{R_i(t) \mid x_i, z_{\mathrm{o},i}, r_i\}$ *may not equal zero; but if* $p(r_i \mid v_i, y_i, \delta_i, \xi)$ *is independent of* y_i *and* δ_i, *then* $E\{R_i(t) \mid x_i, z_{\mathrm{o},i}, r_i\} = 0$;
(iii) *if the missing data are missing at random, then*

$$R_i(t) = N_i(t) - \int_0^t Y_i(u) E\left[\exp\{(x_i^{\mathrm{T}}, z_i^{\mathrm{T}})\beta\} \mid x_i, z_{\mathrm{o},i}, \delta_i, y_i\right] h_0(u)\, \mathrm{d}u$$

and $R_i(t)$ *is independent of* ξ; *moreover, for any* t, $\sum_{i=1}^n \hat{R}_i(t) = 0$;
(iv) $U_1(t \mid \eta) = \sum_{i=1}^n \int_0^t U_{i,1}(u; \eta)\, \mathrm{d}\{R_i(u)\}$;
(v) $U_2(t \mid \eta) \neq \sum_{i=1}^n \int_0^t U_{i,2}(u; \eta)\, \mathrm{d}\{R_i(u)\}$.

Theorem 2 characterizes the behaviour of $R_i(t)$ and $\hat{R}_i(t)$. First, $E\{R_i(t)\}$, $E\{R_i(t) \mid x_i\}$ and $E\{R_i(t) \mid x_i, z_{\mathrm{o},i}\}$ are unbiased, whereas $E\{R_i(t) \mid x_i, z_{\mathrm{o},i}, r_i\}$ is biased. Second, the missing-data indicators can be dropped from $R_i(t)$ under the missing-at-random assumption. Third, the conditional martingale residuals share some properties with ordinary residuals in linear models and martingale residuals in the Cox regression model. Fourth, when there are missing covariates, we cannot replace $N_i(t)$ by $R_i(t)$ in the score residual process.

3·3. *Conditional residual process without incorporating missing data*

We use the conditional martingale residuals to develop test statistics to check model assumptions in the Cox regression model with missing covariates. These statistics are designed to test the null hypothesis $H_0^{(0)} : E\{M(t) \mid x, z\} = 0$ for some η and all t against the alternative $H_1^{(0)} : E\{M(t) \mid x, z\} \neq 0$ for all η and some t. However, since some components of z are missing, we may wish to test the equality $h(t \mid x) = E\{R(t) \mid x\} = 0$ instead; so we test

$$H_0^{(1)} : h(t \mid x) = 0 \text{ for some } \eta \text{ and all } t,$$
$$H_1^{(1)} : h(t \mid x) \neq 0 \text{ for all } \eta \text{ and some } t.$$

Note that $h(t \mid x) = 0$ is only a necessary condition for $E\{M(t) \mid x, z\} = 0$, so accepting $h(t \mid x) = 0$ does not imply acceptance of $H_0^{(0)}$.

We can construct statistics for testing $H_0^{(1)}$ as follows. Using the same reasoning as in Escanciano (2006) and Zhu et al. (2009), we can show that $H_0^{(1)}$ is equivalent to testing $E\{R(t)1(x^{\mathrm{T}}\varphi \leqslant u)\} = 0$ for almost every (φ, u) and all $t \in [0, \tau]$. Thus, we may define a stochastic process

$$I_1(\varphi, u, t; \eta) = n^{-1/2} \sum_{i=1}^{n} 1(x_i^{\mathrm{T}}\varphi \leqslant u) R_i(t),$$

where $(\varphi, u) \in \Pi$ and $t \in [0, \tau]$. We regard $I_1(\varphi, u, t; \eta)$ as a stochastic process indexed by (φ, u, t) and use it to construct a Cramer–von Mises test statistic

$$\mathrm{CM}_1(t) = \int_{\Pi} \left| I_1(\varphi, u, t; \hat{\eta}) \right|^2 F_{n,\varphi}(\mathrm{d}u)\, \mathrm{d}\varphi,$$

where $F_{n,\varphi}(u)$ is the empirical distribution function of $\{x_i^{\mathrm{T}}\varphi : i = 1, \ldots, n\}$. Large values of $\mathrm{CM}_1(t)$ lead to rejection of $H_0^{(1)}$. Compared with other test statistics based on martingale residual processes (Lin et al., 1993), $\mathrm{CM}_1(t)$ avoids both numerical integration in high dimensions and high-dimensional maximization.

THEOREM 3. *Under Assumptions 1–7, $I_1(\varphi, u, t; \hat{\eta})$ is asymptotically equivalent to the sum of $I_1(\varphi, u, t; \eta_*)$ and $n^{1/2}[h_1(\varphi, u, t; \eta_*)^{\mathrm{T}}(\hat{\beta} - \beta_*) + h_2(\varphi, u, t; \eta_*)^{\mathrm{T}}(\hat{\alpha} - \alpha_*) + \int_0^{\tau} h_3(\varphi, u, t; \eta_*)(s)\, \mathrm{d}\{\hat{H}_0(s) - H_0(s)\}]$, where $h_1(\varphi, u, t; \eta_*)$, $h_2(\varphi, u, t; \eta_*)$ and $h_3(\varphi, u, t; \eta_*)(s)$ are defined in the Supplementary Material. Moreover, as $n \to \infty$, $I_1(\varphi, u, t; \hat{\eta})$ converges in distribution to a zero-mean Gaussian process $G_1(\varphi, u, t)$ and $\mathrm{CM}_1(t)$ converges in distribution to $\int_{\Pi} |G_1(\varphi, u, t)|^2 F_{\varphi}(\mathrm{d}u)\, \mathrm{d}\varphi$.*

Theorem 3 characterizes the asymptotic null distributions of $I_1(\varphi, u, t; \hat{\eta})$ and $\mathrm{CM}_1(t)$. Based on this result, we can develop a resampling method to approximate the null distribution of CM_1. Let $\{v_i^{(b)} : i = 1, \ldots, n\}$ be a random sample from the $N(0, 1)$ distribution. We calculate

$$I_1(\varphi, u, t; \hat{\eta})^{(b)} = n^{-1/2} \sum_{i=1}^{n} v_i^{(b)} \left\{ \hat{R}_i(t) 1(x_i^{\mathrm{T}}\varphi \leqslant u) + n \hat{\Delta}_n(\varphi, u, t)^{\mathrm{T}} J_n^{-1} \psi_{n,i} \right\},$$

where $\psi_{n,i}$ denotes the score vector for (β, α) and the $\hat{h}_0(y_i)$ for all uncensored observations, and $\hat{\Delta}_n(\varphi, u, t)$ includes $h_1(\varphi, u, t; \hat{\eta})$, $h_2(\varphi, u, t; \hat{\eta})$ and all $h_3(\varphi, u, t; \hat{\eta})(y_i)$ for $\delta_i = 1$. We then calculate the test statistics $\{\mathrm{CM}_1(t)^{(b)} : b = 1, \ldots, B\}$ and approximate the p-value of $\mathrm{CM}_1(t)$. Theoretically, we can show that this resampling method is asymptotically valid.

COROLLARY 1. *Suppose that Assumptions 1–7 hold. As $n \to \infty$, conditional on the observed data, $I_1(\varphi, u, t; \hat{\eta})^{(q)}$ converges weakly to the same Gaussian process as $I_1(\varphi, u, t; \hat{\eta})$.*

3·4. *Conditional residual process incorporating missing data*

Here we consider using the missing covariates z_i to improve the power of $I_1(\varphi, u, t; \eta)$ in detecting potential model misspecification. Since $1(x^{\mathrm{T}}\varphi \leqslant u)$ in $I_1(\varphi, u, t; \eta)$ does not involve the missing covariates z, we may lose power in detecting the misspecification of $H_0^{(0)}$ in the missing-covariate space. In particular, if the fraction of missing covariates is small, then it is very inefficient to drop all the information in z.

We first suppose that $p(r_i \mid x_i, z_i, y_i, \delta_i; \xi)$ is independent of y_i and δ_i. It can be shown that

$$E\{R_i(t)1(\hat{v}_i^{\mathrm{T}}\tilde{\varphi} \leqslant u) \mid x_i, z_{\mathrm{o},i}\} = 0 \quad (i = 1, \ldots, n),$$

where $(\tilde{\varphi}, u) \in \widetilde{\Pi} = \{\tilde{\varphi} \in \mathbb{R}^{p_1+p_2} : \tilde{\varphi}^{\mathrm{T}}\tilde{\varphi} = 1\} \times [-\infty, \infty]$ and $\hat{v}_i = \{x_i, z_{\mathrm{o},i}, z_{\mathrm{m},i}(\hat{\alpha})\}$. We are thus able to incorporate the additional information from $z_{\mathrm{o},i}$ into the indicator function $1(\hat{v}_i^{\mathrm{T}}\tilde{\varphi} \leqslant t)$. We propose the stochastic process

$$I_2(\tilde{\varphi}, u, t; \eta) = n^{-1/2} \sum_{i=1}^{n} 1(\hat{v}_i^{\mathrm{T}}\tilde{\varphi} \leqslant u) R_i(t).$$

Plotting $I_2(\tilde{\varphi}, u, t; \hat{\eta})$ against t for a specific $\tilde{\varphi}$ provides an exploratory tool for detecting the form of misspecification of assumption (1). Then we introduce the corresponding Cramer–von Mises test statistic based on $I_2(\tilde{\varphi}, u, t; \hat{\eta})$, denoted by CM_2. Large values of CM_2 lead to rejection of the hypothesis that $E\{R_i(t) \mid x_i, z_{\mathrm{o},i}, r_i\} = 0$, which may be caused either by dependence of $p(r_i \mid x_i, z_i, y_i, \delta_i; \xi)$ on (y_i, δ_i) or by $E\{R(t) \mid x, z\} \neq 0$.

Second, we develop a general strategy for incorporating the information from the missing data. We investigate whether we can use the imputed missing covariate data $\hat{v}_i$ when $p(r_i \mid x_i, z_i, y_i, \delta_i, \xi)$ depends on y_i and δ_i. It can be shown that

$$E\{R_i(t)1(\hat{v}_i^{\mathrm{T}}\tilde{\varphi} \leqslant t) \mid x_i, z_{\mathrm{o},i}\} = E[E\{R_i(t) \mid x_i, z_{\mathrm{o},i}, r_i\}1(\hat{v}_i^{\mathrm{T}}\tilde{\varphi} \leqslant t) \mid x_i, z_{\mathrm{o},i}] \neq 0,$$

which arises from the facts that $v_i(\alpha)$ is a function of v_i and r_i and that $E\{R_i(t) \mid x_i, z_{\mathrm{o},i}, r_i\} \neq 0$. We propose to construct a density function for z_i given x_i, denoted by $\hat{p}(z_i \mid x_i)$, using either parametric methods or nonparametric methods based on all the observed data, and then simulate z_i in the space of the missing covariate data for all observations rather than only imputing the missing covariates $z_{\mathrm{m},i}$. For simplicity, we use $p(z_i \mid x_i; \hat{\alpha})$ to simulate $\{z_i^{(b)} : i = 1, \ldots, n\}$ for $b = 1, \ldots, B_3$. Let $v_i^{(b)} = (x_i, z_i^{(b)})$. Then we propose a conditional martingale residual process

$$I_3(\tilde{\varphi}, u, t; \eta)^{(b)} = n^{-1/2} \sum_{i=1}^{n} 1(v_i^{(b)\mathrm{T}}\tilde{\varphi} \leqslant u) R_i(t).$$

We can plot $I_3(\tilde{\varphi}, u, t; \hat{\eta})^{(b)}$ against u for a specific $\tilde{\varphi}$ as an exploratory tool for detecting possible model misspecification. Similar to the above, we can construct a corresponding Cramer–von Mises test statistic based on $I_3(\tilde{\varphi}, u, t; \hat{\eta})^{(b)}$, which we denote by $\mathrm{CM}_3^{(b)}$. Large values of $\mathrm{CM}_3^{(b)}$ lead to rejection of the hypothesis that $E\{R(t) \mid x_i, z_{\mathrm{o},i}\} = 0$.

Following Zhu et al. (2009), we can establish the asymptotic distributions of $I_2(\tilde{\varphi}, u, t; \hat{\eta})$, $I_3(\tilde{\varphi}, u, t; \hat{\eta})^{(b)}$, CM_2 and $\mathrm{CM}_3^{(b)}$. For brevity, we present only the asymptotic null distribution of $I_2(\tilde{\varphi}, u, t; \hat{\eta})$ below.

COROLLARY 2. *If Assumptions 1–8 hold, then $I_2(\tilde{\varphi}, u, t; \hat{\eta})$ converges in distribution to a zero-mean Gaussian process $G_2(\varphi, u, t)$ defined in the Supplementary Material.*

Based on the results in Corollary 2, we can also develop a resampling method to approximate the null distribution of $\mathrm{CM}_2(t)$ in order to calculate the p-values of the test statistic $\mathrm{CM}_2(t)$.

4. Simulation studies

4·1. *Case-deletion measures and martingale residuals*

We generated 100 datasets from a Cox regression model with missing covariates. Each dataset consists of $n = 200$ observations $\{(x_i, z_i, \delta_i, y_i) : i = 1, \ldots, n\}$, with a completely observed covariate x_i and two missing covariates $z_i = (z_{i1}, z_{i2})^{\mathrm{T}}$. The covariate x_i was generated from a Ber(0·5) distribution; conditional on x_i, z_{i1} was generated from the logistic regression model $\mathrm{logit}\{\mathrm{pr}(z_{i1} = 1 \mid x_i, \alpha_1)\} = \alpha_{10} + \alpha_{11} x_i$ with $\alpha_{10} = -1{\cdot}0$ and $\alpha_{11} = -0{\cdot}5$; conditional on (x_i, z_{i1}), z_{i2} was generated from a $N(\alpha_{20} + \alpha_{21} x_{i1} + \alpha_{22} z_{i1},\ \alpha_{23})$ distribution, where $(\alpha_{20}, \alpha_{21}, \alpha_{22}, \alpha_{23}) = (0{\cdot}2, 0{\cdot}1, -0{\cdot}4, 1)$. The survival time T_i was independently generated from $\lambda(t \mid c_i; \beta) = h_0(t) \exp(x_i \beta_1 + z_{i1} \beta_2 + z_{i2} \beta_3)$ with $h_0(t) = 0{\cdot}56$ and $\beta = (0{\cdot}5, 0{\cdot}5, -1{\cdot}0)^{\mathrm{T}}$, and the censoring times C_i were independently generated from a Un(0, 3) distribution. We then let $y_i = \min(T_i, C_i)$ and set $\delta_i = 1$ when $T_i \leqslant C_i$ and 0 otherwise. The missing data z_{i1} were generated from the logistic regression model $\mathrm{logit}\{\mathrm{pr}(r_{i1} = 1 \mid y_i, c_i; \xi_1)\} = \xi_{10} + \xi_{11} y_i + \xi_{12} x_i + \xi_{13} z_{i2}$ with $\xi_1 = (0{\cdot}5, 0{\cdot}3, 0{\cdot}5, -0{\cdot}5)^{\mathrm{T}}$, and the missing data z_{i2} were generated from the logistic regression model $\mathrm{logit}\{\mathrm{pr}(r_{i2} = 1 \mid y_i, r_{i1}, x_i, z_i, \xi_2)\} = \xi_{20} + \xi_{21} y_i + \xi_{22} r_{i1} + \xi_{23} x_i + \xi_{24} z_{i1}$ with $\xi_2 = (0{\cdot}3, 0{\cdot}3, 0{\cdot}4, -0{\cdot}2, 0{\cdot}2)^{\mathrm{T}}$. In the above simulation design, each simulated dataset has about 44% censored values of the y_i, about 23% missing covariates z_{i1}, and about 37% missing covariates z_{i2}.

We investigate the performance of different diagnostic measures on the simulated datasets. Two outliers are introduced into each simulated dataset. In the first dataset, we perturbed the 41st observation by adding s to each survival time, i.e., $y_i + s$, and perturbed the 90th observation by adding s to $z_{i,2}$, i.e., $z_{i,2} + s$. For each of the other 99 datasets, we selected the two observations closest to the 41st and 90th observations according to the values of $(y_i, \delta_i, x_i, z_i, r_i)$ and perturbed these two observations in the same way as in the first dataset. We varied the value of s to represent different degrees of perturbation. We fitted the same missing-not-at-random model used to generate the simulated datasets, and then calculated various diagnostic measures and their detection probabilities. Table 1 summarizes the detection probabilities of various diagnostic measures for the 100 simulated datasets. As the degree of perturbation increases, the detection probabilities increase in magnitude. Lowering the threshold for the detection probabilities from 97·5% to 90% increases the probability of detecting the induced outliers. Overall, our detection probability is effective for detecting outliers.

4·2. *Cramer–von Mises goodness-of-fit test statistics*

The goal of this simulation is to assess the empirical performance of $\mathrm{CM}_1(\tau)$ and $\mathrm{CM}_2(\tau)$ and their associated resampling method. We generated datasets from a Cox regression model with two completely observed covariates $x_i = (x_{i1}, x_{i2})^{\mathrm{T}}$ and one missing covariate z_i as follows. In this simulation study, x_{i1} and x_{i2} were independently generated from the normal distributions $N(0, 1)$ and $N(0, 0{\cdot}5^2)$, respectively; conditional on x_i, z_i was generated from a normal distribution $N(\alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2},\ \alpha_3)$, where $(\alpha_0, \alpha_1, \alpha_2, \alpha_3) = (0{\cdot}5, 0{\cdot}1, -0{\cdot}4, 1)$. The survival times T_i were independently generated from $\lambda(t \mid x_i, z_i; \beta) = h_0(t) \exp(x_{i1} \beta_1 + x_{i2} \beta_2 + z_i \beta_3 + c x_{i1}^2)$ with $h_0(t) = 1{\cdot}0$ and $\beta = (1{\cdot}0, 1{\cdot}0, -1{\cdot}0)^{\mathrm{T}}$, and the censoring times C_i were independently generated from a Un(0, 11) distribution. We then set $y_i = \min(T_i, C_i)$ and set $\delta_i = 1$ when $T_i \leqslant C_i$ and 0 otherwise. The missing data z_i were assumed to be missing at random and generated from the logistic regression model $\mathrm{logit}\{\mathrm{pr}(r_i = 1 \mid x_i; \xi)\} = \xi_0 + \xi_1 x_{i1} + \xi_2 x_{i2}$. We considered two sets of ξ: (I) $\xi = (1{\cdot}085, 0{\cdot}2, 0{\cdot}1)^{\mathrm{T}}$; (II) $\xi = (-0{\cdot}015, 0{\cdot}2, 0{\cdot}1)^{\mathrm{T}}$. The averages and ranges of the missing-data fractions are, respectively, 25% and (19·2%, 31·6%) under (I) and 50% and (44·0%, 58·4%) under (II). We considered $c = 0,\ 0{\cdot}25, 0{\cdot}50$ and $0{\cdot}75$. The average

Table 1. *Summary of detection probabilities* (%) *of* $\mathrm{QD}_{i,1}$, $\mathrm{QD}_{i,2}$, $\mathrm{QD}_{i,3}$, $\mathrm{QD}_{i,h_0(\cdot)}$, *the martingale residual* $\hat{R}_i$, *and the score residual* s_{i1} *for* 100 *simulated datasets under the missing-not-at-random model*

$y_{41}+s$		$\mathrm{pQD}_{i,1}$	$\mathrm{pQD}_{i,2}$	$\mathrm{pQD}_{i,3}$	$\mathrm{pQD}_{i,h_0(\cdot)}$	$\mathrm{p}\hat{R}_i$	$\mathrm{p}s_{i1}$	Max
$s=0{\cdot}1$	Median	83	97	38	16·5	55·5	95	98
	Q1	32·5	94·5	30	6	23	70·5	96
	Q3	98	100	49	40·5	95	99	100
	$\geqslant$90%	44	95	0	0	33	56	96
	$\geqslant$97·5%	26	48	0	0	15	39	62
$s=0{\cdot}2$	Median	95	97	36	29	92	98·5	99
	Q1	78·5	94	23·5	13	48	89·5	98
	Q3	99	99	47·5	56	98	100	100
	$\geqslant$90%	66	97	0	0	53	75	100
	$\geqslant$97·5%	41	48	0	0	34	57	76
$z_{90,2}+s$		$\mathrm{pQD}_{i,1}$	$\mathrm{pQD}_{i,2}$	$\mathrm{pQD}_{i,3}$	$\mathrm{pQD}_{i,h_0(\cdot)}$	$\mathrm{p}\hat{R}_i$	$\mathrm{p}s_{i1}$	Max
$s=1{\cdot}5$	Median	95	85	55	36·5	79	99	99
	Q1	90	80	42·5	26·5	73·5	97·5	97·5
	Q3	97	92	66	46	83	99	99
	$\geqslant$90%	78	33	1	0	5	100	100
	$\geqslant$97·5%	24	3	1	0	1	75	75
$s=2{\cdot}5$	Median	99	99	85	41	86	100	100
	Q1	98	98	75	31	82·5	100	100
	Q3	100	100	93	52	89	100	100
	$\geqslant$90%	100	99	41	0	24	100	100
	$\geqslant$97·5%	80	87	14	0	0	100	100

Q1, Median and Q3, the 25th, 50th and 75th percentiles of 100 detection probabilities; Max, maximum value calculated as $\max(\mathrm{pQD}_{i,1}, \mathrm{pQD}_{i,2}, \mathrm{pQD}_{i,3}, \mathrm{pQD}_{i,h_0(\cdot)}, \mathrm{p}\hat{R}_i, \mathrm{p}s_{i1})$.

Table 2. *Rejection rates* (%) *of* $\mathrm{CM}_1(\tau)$ *and* $\mathrm{CM}_2(\tau)$ *at the* 5% *significance level*

	Average missing data fraction							
	25%				50%			
	Complete-case analysis		Analysis of all cases		Complete-case analysis		Analysis of all cases	
c	$\mathrm{CM}_1(\tau)$	$\mathrm{CM}_2(\tau)$	$\mathrm{CM}_1(\tau)$	$\mathrm{CM}_2(\tau)$	$\mathrm{CM}_1(\tau)$	$\mathrm{CM}_2(\tau)$	$\mathrm{CM}_1(\tau)$	$\mathrm{CM}_2(\tau)$
0·00	3	2	7	5	2	2	4	4
0·25	67	42	77	57	46	23	72	41
0·50	99	94	94	92	92	83	87	80
0·75	100	100	96	95	98	93	88	86

censoring percentages are, respectively, 24·6%, 21·4%, 18·7% and 16·7% for $c=0$, 0·25, 0·50 and 0·75. For each combination of c and ξ, we generated 100 datasets.

For all simulated datasets, we fitted the Cox regression model (1) with $\lambda(t \mid x_i, z_i, \beta) = h_0(t)\exp(x_{i1}\beta_1 + x_{i2}\beta_2 + z_i\beta_3)$, assuming missingness at random. We carried out the complete-case analysis and the all-case analysis. Thus, the model would be misspecified if $c \neq 0$, and the misspecification would be due to the covariate x_{i1}^2. We set $B=500$ to calculate the p-values of all test statistics. The significance level was fixed at 0·05.

The rejection rates are presented in Table 2. As expected, the power of both $\mathrm{CM}_1(\tau)$ and $\mathrm{CM}_2(\tau)$ to detect model misspecification increases with c and decreases with the missing-data fraction. Moreover, the power of $\mathrm{CM}_1(\tau)$ is higher than that of $\mathrm{CM}_2(\tau)$, but $\mathrm{CM}_1(\tau)$ has slightly greater Type I errors than $\mathrm{CM}_2(\tau)$ when the missing-data fraction is low. For the complete-case analysis,

the Type I error rates of CM_1 and CM_2 are close to 0·025. In contrast, for the all-case analysis, the Type I error rates of CM_1 and CM_2 are 0·07 and 0·05 when the average missing-data fraction is 25% and are 0·04 and 0·04 when the average missing-data fraction is 50%. When $c = 0{\cdot}25$, the all-case analysis outperforms the complete-case analysis in terms of detecting model misspecification. However, this is not the case when $c \geqslant 0{\cdot}5$, which may be due to the robustness of the complete-case analysis when the data are truly missing at random.

5. Analysis of lung cancer data

We analyse data from a Phase III advanced non-small-cell lung cancer clinical trial conducted at the University of North Carolina at Chapel Hill (Socinski et al., 2002). The goal of this trial was to compare a defined duration of therapy with continuous therapy followed by second-line therapy in order to determine the optimal duration of therapy for non-small-cell lung cancer patients. The study involved $n = 230$ patients. We consider five prognostic factors: $x_1 =$ treatment, which takes the value 1 if the subject received a defined duration of therapy and 0 otherwise; $x_2 =$ gender, which equals 1 if the subject is male and 0 otherwise; $x_3 =$ age in years; $z_1 =$ Apex, which equals 1 if the tumour was at the top of the lung and 0 otherwise; and $z_2 =$ FACT-G score. Of these five prognostic factors, z_1 and z_2 had missing information while x_1, x_2 and x_3 were completely observed for all cases. In this dataset, 52·74% of the subjects had missing values in at least one of z_1 and z_2. The outcome variable is time to disease progression, which is continuous and subject to right censoring; the censoring indicator δ_i is equal to 1 if the ith subject showed disease progression and 0 otherwise. The median follow-up time is 3·94 months, and the range of the follow-up time is (0·1, 27·61) months. A summary of the dataset can be found in Chen et al. (2009).

We fitted the Cox regression model (1) to the data, where $v_i = (x_i^{\mathrm{T}}, z_i^{\mathrm{T}})^{\mathrm{T}}$ and $\beta = (\beta_1, \ldots, \beta_5)^{\mathrm{T}}$ with $p_1 = 3$ and $p_2 = 2$. We model two missing covariates z_i conditional on x_i as $p(z_{i1} \mid x_i; \alpha)\, p(z_{i2} \mid x_i, z_{i1}; \alpha)$. We use a logistic regression model for z_{i1} and a normal linear regression model for z_{i2}. Specifically, we have $\text{logit}\{p(z_{i1} \mid x_i; \alpha)\} = z_{i1}(\alpha_{10} + \sum_{k=1}^{3} \alpha_{1k} x_{ik})$ and $z_{i2} \sim N(\alpha_{20} + \sum_{k=1}^{3} \alpha_{2k} x_{ik} + \alpha_{24} z_{i1},\ \alpha_{25})$, where $\alpha_1 = (\alpha_{10}, \alpha_{11}, \alpha_{12}, \alpha_{13})^{\mathrm{T}}$ and $\alpha_2 = (\alpha_{20}, \ldots, \alpha_{25})^{\mathrm{T}}$. We consider both missing-at-random and missing-not-at-random models for r_i. Under the missing-not-at-random model, we take $p(r_i \mid v_i, y_i, \delta_i; \xi) = p(r_{i1} \mid v_i, y_i, \delta_i; \xi_1)\, p(r_{i2} \mid r_{i1}, v_i, y_i, \delta_i; \xi_2)$ with $\xi = (\xi_1^{\mathrm{T}}, \xi_2^{\mathrm{T}})^{\mathrm{T}}$. Moreover, logistic regression models are specified for $p(r_{i1} \mid v_i, y_i, \delta_i; \xi_1)$ and $p(r_{i2} \mid r_{i1}, v_i, y_i, \delta_i; \xi_2)$, where ξ_1 and ξ_2 are the vectors of the corresponding regression coefficients. Under the missing-at-random model, $p(r_i \mid v_i, y_i, \delta_i; \xi) = p(r_i \mid x_i, y_i, \delta_i; \xi)$ and a logistic regression model is specified for $p(r_i \mid x_i, y_i, \delta_i; \xi)$. For comparison, we also consider the complete-case analysis.

Table 3 shows the maximum partial likelihood estimate of β for the complete-case analysis and the maximum likelihood estimates of β under the missing-at-random and missing-not-at-random models for the missing-data mechanism. We can see some differences between the estimates in Table 3. In the complete-case analysis, the p-value for β_1 is 0·062 while that for β_4 is 0·032. Hence, in the complete-case analysis, treatment is not significant but Apex is significant at the 0·05 significance level. However, the p-values are 0·006 and 0·006 for β_1 and 0·016 and 0·015 for β_4 under the missing-at-random and missing-not-at-random models, respectively, implying that treatment and Apex may be significantly associated with time to disease progression. Therefore, in terms of time to disease progression, continuous therapy followed by second-line therapy may be more beneficial than a defined duration of therapy, based on the analysis incorporating all of the cases. Also, the standard errors obtained from the analysis incorporating all of the cases are consistently smaller than those from the complete-case analysis for all of the β_j. In addition, the

Table 3. *Maximum likelihood estimates of β based on complete-case, missing-at-random and missing-not-at-random analyses of the lung cancer data. For each β_k, the efficiency shown in the last column represents the ratio of the standard error of $\hat{\beta}_k$ for the complete-case analysis to that for the missing-at-random (or missing-not-at-random) analysis*

Model	Parameter	Estimate	SE	Z-statistic	p-value	95% CI	Efficiency
Complete case	β_1	0·47	0·25	1·86	0·06	(−0·02, 0·97)	1·00
	β_2	0·07	0·24	0·28	0·78	(−0·41, 0·55)	1·00
	β_3	−0·02	0·13	−0·15	0·88	(−0·28, 0·24)	1·00
	β_4	0·88	0·41	2·14	0·03	(0·07, 1.68)	1·00
	β_5	−0·14	0·12	−1·16	0·25	(−0·37, 0·10)	1·00
Missing at random	β_1	0·48	0·18	2·72	0·01	(0·13, 0·82)	1·45
	β_2	0·17	0·18	0·97	0·33	(−0·18, 0·53)	1·35
	β_3	−0·02	0·09	−0·24	0·81	(−0·20, 0·16)	1·44
	β_4	0·91	0·38	2·40	0·02	(0·17, 1·66)	1·08
	β_5	−0·05	0·11	−0·49	0·62	(−0·26, 0·16)	1·12
Missing not at random	β_1	0·48	0·18	2·73	0·01	(0·14, 0·82)	1·45
	β_2	0·17	0·18	0·96	0·34	(−0·18, 0·53)	1·35
	β_3	−0·02	0·09	−0·24	0·81	(−0·20, 0·16)	1·44
	β_4	0·92	0·38	2·43	0·015	(0·18, 1·67)	1·08
	β_5	−0·05	0·11	−0·48	0·63	(−0·26, 0·16)	1·12

SE, standard errors; CI, confidence interval.

two sets of maximum likelihood estimates of β are very similar. Under the missing-not-at-random model, the p-values for the coefficients associated with z_i in $p(r_i \mid v_i, y_i, \delta_i; \xi)$ are greater than 0·34, which could suggest that there is no evidence against the missing-at-random assumption.

We calculated the test statistics CM_1 and CM_2 to be 0·377 and 0·637, respectively. By setting $B = 1000$, we approximated the p-values of CM_1 and CM_2 by 0·303 and 0·127, respectively. These results may also indicate that $E\{R_i(t) \mid x_i\} \neq 0$ or $E\{R_i(t) \mid x_i, z_i\} \neq 0$, and $p(r_i \mid x_i, z_i, y_i, \delta_i; \xi)$ does not depend on (y_i, δ_i).

Figure 1 plots the detection probabilities of selected diagnostic measures under the missing-at-random and missing-not-at-random models. Additional results are shown in the Supplementary Material. The purpose of plotting the detection probabilities corresponding to $\text{QD}_{i,1}$, $\text{QD}_{i,2}$, $\text{QD}_{i,3}$, and $\text{QD}_{i,h_0(\cdot)}$ is to identify influential observations due to the specifications of the regression component of the Cox model, the covariate model, the logistic regression models for the missing-data binary indicators, and the baseline hazard component of the Cox model, respectively. In addition, the plots corresponding to $\hat{R}_i$ are used to determine the appropriateness of the entire Cox model, while the plots corresponding to $\hat{s}_{i3}$ and $\hat{s}_{i5}$ are used to check the proportional hazard assumptions for age and FACT-G score, respectively. For the 111 complete cases, both the missing-at-random and the missing-not-at-random models detected the same 13 outlying cases with maximum detection probabilities greater than 0·95. Of the 119 subjects who had at least one missing value in Apex or FACT-G score, the same 12 subjects had maximum detection probabilities greater than 0·95 under both the missing-at-random and the missing-not-at-random models, six subjects had maximum detection probabilities greater than 0·95 only under the missing-at-random model, and four subjects had maximum detection probabilities greater than 0·95 only under the missing-not-at-random model. For the six missing-at-random outlying cases, the maximum detection probabilities range from 0·902 to 0·932 under the missing-not-at-random model and range from 0·967 to 0·992 under the missing-at-random model. For the four missing-not-at-random outlying cases, the maximum detection probabilities are 0·80, 0·58, 0·825 and 0·898 under the missing-at-random model and 0·96, 0·958, 0·984 and 0·990 under the

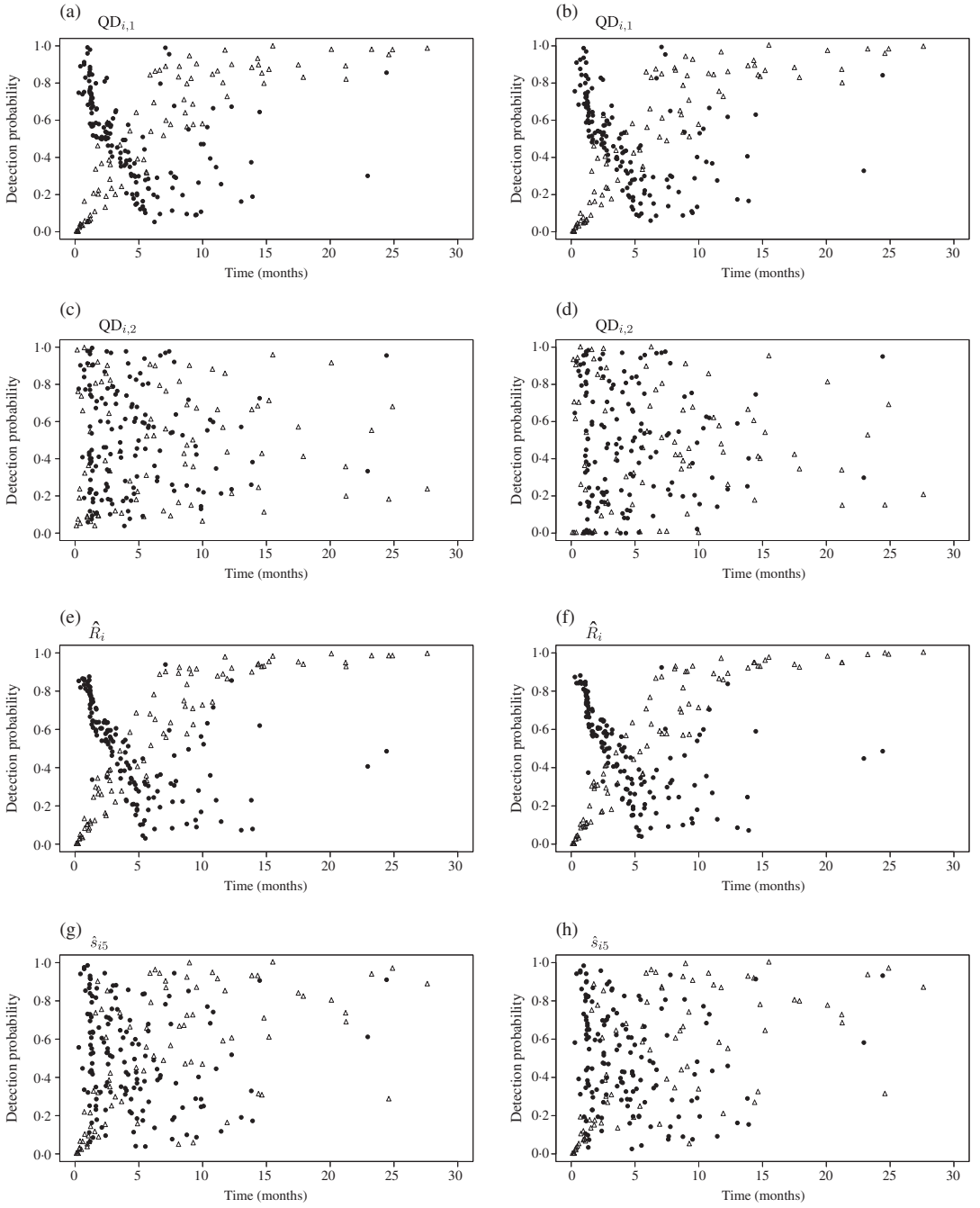

Fig. 1. Plots of detection probabilities of $QD_{i,1}$, $QD_{i,2}$, the conditional martingale residuals $\hat{R}_i$, and the score residuals $\hat{s}_{i5}$ for the missing-at-random (panels (a), (c), (e), (g)) and missing-not-at-random (panels (b), (d), (f), (h)) analyses of the lung cancer data. Filled circles represent the detection probabilities for progression subjects, and empty triangles represent the detection probabilities for censored subjects.

missing-not-at-random model. The disagreements between the missing-at-random and missing-not-at-random models for these four cases were in the values and detection probabilities of $QD_{i,2}$. Overall, the detection probabilities under the missing-at-random model are very close to those under the missing-not-at-random model. The outlying case with the greatest maximum detection

probabilities is the subject whose FACT-G score is 34, which is the smallest value among all subjects, with mean FACT-G score 78·14. In this case, the values of s_{i5} are 5·77, 4·84 and 4·82, and the corresponding detection probabilities are all 1·0 under the complete-case analysis and under the missing-at-random and missing-not-at-random models.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes details of the semi-bootstrap method, proofs of the theoretical results, additional simulations, real-data analysis results, the lung cancer data used in § 5, and the computer code.

ACKNOWLEDGEMENT

We thank the editor, associate editor and two referees for valuable suggestions, which have greatly helped to improve the presentation, as well as Dr Dongling Zeng for helpful discussions. This research was supported by the U.S. National Institutes of Health and National Science Foundation.

REFERENCES

ANDREWS, D. W. K. (1994). Empirical process methods in econometrics. In *Handbook of Econometrics*, vol. 4. R. F. Engle & D. L. McFadden, eds. Amsterdam: Elsevier, pp. 2248–92.

BARLOW, W. E. (1997). Global measures of local influence for proportional hazards regression models. *Biometrics* **53**, 1157–62.

BARLOW, W. E. & PRENTICE, R. L. (1988). Residuals for relative risk regression. *Biometrika* **75**, 65–74.

CHEN, H. Y. & LITTLE, R. J. (1999). Proportional hazards regression with missing covariate. *J. Am. Statist. Assoc.* **94**, 896–908.

CHEN, M., IBRAHIM, J. G. & SHAO, Q. M. (2009). Maximum likelihood inference for the Cox regression model with applications to missing covariates. *J. Mult. Anal.* **100**, 2018–30.

CHEN, M. H., SHAO, Q. M. & IBRAHIM, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer.

COOK, R. D. (1986). Assessment of local influence (with Discussion). *J. R. Statist. Soc.* B **48**, 133–69.

COOK, R. D. & WEISBERG, S. (1982). *Residuals and Influence in Regression*. Boca Raton, Florida: Chapman & Hall.

COPAS, J. B. & EGUCHI, S. (2005). Local model uncertainty and incomplete-data bias (with Discussion). *J. R. Statist. Soc.* B **67**, 459–513.

COX, D. R. (1972). Regression models and life-tables (with Discussion). *J. R. Statist. Soc.* B **34**, 187–220.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–76.

COX, D. R. & SNELL, E. J. (1968). A general definition of residuals (with Discussion). *J. R. Statist. Soc.* B **30**, 248–75.

DANIELS, M. J. & HOGAN, J. W. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Boca Raton, Florida: Chapman & Hall.

ESCANCIANO, J. C. (2006). A consistent diagnostic test for regression models using projection. *Economet. Theory* **22**, 1030–51.

ESCOBAR, L. A. & MEEKER, W. Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics* **48**, 507–28.

HENDERSON, R. & OMAN, P. (1993). Inference in linear hazard models. *Scand. J. Statist.* **20**, 195–212.

HERRING, A. & IBRAHIM, J. G. (2001). Likelihood-based methods for missing covariates in the Cox proportional hazards model. *J. Am. Statist. Assoc.* **96**, 292–302.

IBRAHIM, J. G., LIPSITZ, S. R. & CHEN, M. (1999). Missing covariates in generalized linear models when the missing data mechanism is nonignorable. *J. R. Statist. Soc.* B **61**, 173–90.

JANSEN, I., MOLENBERGHS, G., AERTS, M., THIJS, H. & VAN STEEN, K. (2003). A local influence approach to binary data from a psychiatric study. *Biometrics* **59**, 410–9.

KLEIN, J. P. & MOESCHBERGER, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.

KOSOROK, M. R. (2007). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer.

LIN, D. Y., WEI, L. J. & YING, Z. L. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–72.

Lipsitz, S. R. & Ibrahim, J. G. (1996). A conditional model for incomplete covariates in parametric regression models. *Biometrika* **83**, 916–22.

Martinussen, T. & Scheike, T. H. (2006). *Dynamic Regression Models for Survival Data*. New York: Springer.

Marzec, L. & Marzec, P. (1997). Generalized martingale-residual processes for goodness-of-fit inference in Cox's type regression models. *Ann. Statist.* **25**, 683–714.

Ossiander, M. (1987). A central limit theorem under metric entropy with bracketing. *Ann. Prob.* **15**, 897–919.

Pettitt, A. N. & Daud, I. B. (1989). Case-weighted measures of influence for proportional hazards regression. *Appl. Statist.* **38**, 51–67.

Pollard, D. (1990). *Empirical Processes: Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 2. Hayward, California: Institute of Mathematical Statistics & Alexandria, Virginia: American Statistical Association.

Scheike, T., Martinussen, T. & Silver, J. (2010). Estimating haplotype effects for survival data. *Biometrics* **66**, 705–15.

Socinski, M. A., Schell, M. J., Peterman, A., Bakri, K., Yates, S., Gitten, R., Unger, P., Lee, J., Lee, J. H., Tynan, M., et al. (2002). Phase III trial comparing defined duration of therapy versus continuous therapy followed by second-line therapy in advanced-stage IIIB/IV non-small-cell lung cancer. *J. Clin. Oncol.* **20**, 1335–43.

Storer, B. E. & Crowley, J. (1985). A diagnostic for Cox regression and general conditional likelihoods. *J. Am. Statist. Assoc.* **80**, 139–47.

Therneau, T. M., Grambsch, P. M. & Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77**, 147–60.

Troxel, A. B., Ma, G. & Heitjan, D. F. (2004). An index of local sensitivity to nonignorability. *Statist. Sinica* **14**, 1221–37.

van der Vaart, A. W. & Wellner, J. A. (1996) *Weak Convergence and Empirical Processes*. New York: Springer.

Verbeke, G., Molenberghs, G., Thijs, H., Lasaffre, E. & Kenward, M. G. (2001). Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics* **57**, 43–50.

Zhu, H. T., Lee, S. Y., Wei, B. C. & Zhou, J. (2001). Case-deletion measures for models with incomplete data. *Biometrika* **88**, 727–37.

Zhu, H. T., Ibrahim, J. G. & Shi, X. Y. (2009). Diagnostic measures for generalized linear models with missing covariates. *Scand. J. Statist.* **36**, 686–712.

[*Received August* 2013. *Revised July* 2015]