Published in final edited form as:

Stat Med. 2009 May 1; 28(10): 1498–1511. doi:10.1002/sim.3557.

Sensitivity analysis to investigate the impact of a missing covariate on survival analyses using cancer registry data

Brian L. Egleston, Ph.D. and Yu-Ning Wong, M.D.

Division of Population Sciences, Fox Chase Cancer Center, Philadelphia, Pennsylvania

Abstract

Having substantial missing data is a common problem in administrative and cancer registry data. We propose a sensitivity analysis to evaluate the impact of a covariate that is potentially missing not at random in survival analyses using Weibull proportional hazards regressions. We apply the method to an investigation of the impact of missing grade on post-surgical mortality outcomes in individuals with metastatic kidney cancer. Data came from the Surveillance Epidemiology and End Results (SEER) registry which provides population based information on those undergoing cytoreductive nephrectomy. Tumor grade is an important component of risk stratification for patients with both localized and metastatic kidney cancer. Many individuals in SEER with metastatic kidney cancer are missing tumor grade information. We found that surgery was protective, but that the magnitude of the effect depended on assumptions about the relationship of grade with missingness.

2 Introduction

The relationship of time-to-event outcomes, such as time to death, with baseline covariates is often of interest to researchers. Cancer researchers are commonly interested in the relationship of survival with tumor characteristics, demographics, or comorbidities. One frequently used data source for such investigations is the Surveillance Epidemiology and End Results (SEER) database. SEER is maintained by the National Cancer Institute and collects tumor characteristics and demographics information about incident cancers in approximately 14% of the United States. SEER data can be linked to Medicare claims to find additional information on treatment and comorbidities. Many studies using SEER data have been published. Berndt *et al.* investigated the relationship of race with survival in individuals with renal cell carcinoma using linked SEER-Medicare data [1]. Also using SEER-Medicare data, Wong *et al.* [2] investigated the impact of treatment with radiation therapy or prostatectomy on mortality in men with localized prostate cancer.

In observational and administrative datasets such as SEER and Medicare, there can be a high level of missing data since information generally comes from medical records collected for clinical and administrative purposes rather than research purposes. Shahinian *et al.* [3] for example, had to exclude over 30,000 out of 92,474 observations because of incomplete medical claims information in a study investigating the risk of fracture after androgen deprivation for prostate cancer.

In this paper, we propose a missing data sensitivity analysis methodology for use with a Weibull Proportional Hazards Regression. This method allows researchers to assess how failure to account for a covariate that is potentially missing not at random might affect inferences. We use a sensitivity analysis paradigm in which identifiability of a model is

induced by fixing key parameters and then estimating effects over a range of the fixed parameters. This approach differs from the sensitivity analysis for survival models of Herring *et al.* [4] in that we only need to specify a model for the missing data mechanism, and not a model for the missing data.

The motivating example for this paper comes from SEER. We examine post-surgical mortality outcomes across cancer grades among patients with metastatic renal cell carcinoma (RCC). Much of the grade data among those with metastatic kidney cancer in the SEER database is missing. Until recently, treatment options for patients with metastatic RCC have been limited, resulting in a median survival of less than one year. Unlike cancer in other organs, removal of the cancerous kidney (cytoreductive nephrectomy) followed by systemic cytokine therapy with interferon or interleukin has been shown to improve median survival compared to patients treated with interferon alone (11.1 vs 8.1 months [5]). Patients who did not undergo a nephrectomy were thought to have poor outcomes [6], and therefore nephrectomy was required for most clinical trials involving the role of systemic cytokines. Three new agents (sorafenib, sunitinib, and temsirolimus) have shown activity in metastatic RCC [7][8][9]. However, it is not known if cytoreductive nephrectomy is still beneficial [10][11]. Clarifying the natural history of these patients could be helpful in designing clinical trials and better estimating the benefit of cytoreductive nephrectomy in the presence of these new agents.

Higher tumor grade is associated with higher cancer specific mortality [12][13]. Therefore, it would be helpful to know if patients with higher grade tumors derive significantly more benefit from surgery than patients with low grade tumors. If this is true, patients with low grade tumors could be spared the morbidity associated with surgery, while patients with higher grade tumors may be offered surgery if there is an established benefit. This may also prompt clinicians to recommend pre-operative biopsies to help direct treatment.

Inferences from analyses in which those with missing tumor data are excluded can potentially give biased inferences about population level effects. For example, if the outcomes of those with missing tumor grade are generally much worse than those with observed tumor grade, and those with high grade tumors are more likely to be missing, then we might underestimate the negative impact of higher tumor grade on outcomes if we only use the observed data in estimation. Since we have no information on missing data, the missing data mechanism is not identifiable. We cannot identify and hence estimate models of interest without making assumptions about the missing data mechanism.

Researchers analyzing data with missing observations, such as the grade data of interest here, have a number of options to identify the missing data mechanism and hence estimate consistent population-level effects of interest. One option is to assume that the reason for missingness is akin to a weighted coin toss that has no relation to any variables in the dataset. This is a missing completely at random (MCAR) assumption [14]. MCAR allows for the estimation of unbiased effects using only those with complete data, although with some loss of efficiency.

Another assumption researchers commonly make is that the reason for missingness is related to observed variables in a sample, often called missing at random (MAR). In cases of MAR, point estimators based on excluding all of the missing data might be biased as well as inefficient. Multiple imputation is a common method of accounting for missing data that is MAR [15]. When missing data are sparse, incorrectly making an MAR assumption might not substantially impact inferences. As the amount of missing data increases, however, the potential for substantial bias due to data missing not at random grows.

In our example, we discuss how varying assumptions about the missingness mechanism can impact parameter estimates in a Weibull proportional hazards regression. In such models, if the reason for missingness is only related to observed and unobserved covariates in the model, then consistent estimates can be obtained by using the fully observed data. However, if the reason for missingness is related to both survival time and unobserved covariates, then the population level Weibull model is not identifiable from the observed data, and additional assumptions are needed for identification. Robins and colleagues [16], and Rathouz [17] discuss identifiability under various missing data mechanisms. In this paper, we consider cases when neither the missing data mechanism nor the Weibull model can be identified using the observed data.

In cases in which there are substantial missing data, researchers might be particularly concerned about how missing data might affect inferences. Stringent assumptions, such as MAR, do not provide the only solution to accounting for missing data. In this paper, we demonstrate how one can use sensitivity analyses to investigate how non-random missingness might affect inferences in survival analyses. If inferences from the sensitivity analysis do not change when compared to inferences based on the complete case data, then researchers can be reassured that missing data are not biasing their analyses. If inferences are dependent on the value of a sensitivity parameter, then researchers can use scientific knowledge to bound the range of the parameter. If such bounding is not scientifically reasonable, then researchers will have a better understanding of how missing data in a particular dataset hinder their ability to make inferences.

This paper is organized as follows. In Section 3, we discuss the data structure. In Section 4 we discuss models of interest and their identifiability. In Section 5, we present our estimation methods. We present a simulation study in Section 6 and then present the data example in Section 7. We conclude the paper with a discussion. The R programs used for analyses are available from the authors.

3 Data Structure and Notation

For this study, let Y_i define time from diagnosis to death as specified in the SEER database for $i = \{1, \dots, n\}$ participants in the dataset. Similarly, let C_i define the time until censoring occurs. $T_i = \min(C_i, Y_i)$. Let D_i be the mortality indicator; $D_i = 1$ if $T_i = Y_i$, 0 otherwise. Let Z_i be a surgical treatment indicator ($Z_i=1$ for surgical treatment, 0 otherwise). Let $\mathbf{X}_i = \{\mathbf{X}_{1i}, X_{2i}\}$ be a set of covariates that could confound the relationship between treatment and mortality time, where $\mathbf{X}_{1i}$ is completely observed and X_{2i} denotes tumor grade and contains missing data. To be consistent with clinical labeling practice, $X_{2i} = 1$ if a tumor is well differentiated (best grade), 2 if a tumor is moderately differentiated, and 3 if a tumor is poorly differentiated (worst grade). In this example, none of the covariates in $\mathbf{X}_i$ are considered to be time-varying. Let M_i be the missing data indicator for baseline data ($M_i=1$ if X_{2i} is missing, 0 otherwise). The observed data are hence,

$$\begin{aligned} O_i &= \{T_i, D_i, Z_i, \mathbf{X}_{1i}, M_i, X_{2i}\} \quad \text{if } M_i=0 \\ O_i &= \{T_i, D_i, Z_i, \mathbf{X}_{1i}, M_i\} \quad \text{if } M_i=1 \end{aligned}$$

4 Identification and Models

In our identification of effects, we make the typical assumption of the Weibull proportional hazards model that the censoring mechanism is independent of the mortality mechanism. This seems to be a reasonable assumption for use with SEER data since date of death information is readily available to the SEER registries. The typical reason why censoring

might occur in the SEER population is that a person is still alive; their death information will be recorded by the registry a short time after they die.

The second major assumption for identification relates to the missing data mechanism. We assume that the following model holds.

Missingness Model

$$\text{logit } P[M_i=1|D_i, \mathbf{X}_i, T_i, Z_i] = h(D_i, \mathbf{X}_{1i}, T_i, Z_i; \boldsymbol{\alpha}^*) + \tau_1 X_{2i} + \tau_2 Z_i X_{2i}$$

where $h(\cdot)$ is a flexible function of the data parameterized by $\boldsymbol{\alpha}^*$. Let $\text{expit}(q) \equiv \exp(q)/(1 + \exp(q))$ and let $\pi(D_i, \mathbf{X}_i, T_i, Z_i; \boldsymbol{\alpha}^*, \boldsymbol{\tau}) = \text{expit}(h(D_i, \mathbf{X}_{1i}, T_i, Z_i; \boldsymbol{\alpha}^*) + \tau_1 X_{2i} + \tau_2 Z_i X_{2i})$ represent the modeled probability of missingness. As part of the sensitivity analysis, we estimate $\boldsymbol{\alpha}^*$ for fixed values of $\boldsymbol{\tau} = \{\tau_1, \tau_2\}$ using the estimating equations described below. This is similar to the sensitivity analyses of Rotnitzky *et al.* [18], Scharfstein *et al.* [19], and Baker *et al.* [20] who advocate identifying similar nonresponse models by the use of such sensitivity parameters. We use the estimates of the inverse probability of observing the data for fixed values of τ_1 and τ_2 as weights in the estimation of the following Weibull proportional hazards model with scale parameter λ^* , shape parameter γ^* , and hazard function $h(t|\mathbf{X}_i, Z_i)$.

Survival Model

$$h(t|\mathbf{X}_i, Z_i) = \exp(j(\mathbf{X}_i, Z_i; \boldsymbol{\beta}^*)) \lambda^* \gamma^* t^{\gamma^*-1}$$

where $j(\cdot)$ is a flexible function of the data parameterized by $\boldsymbol{\beta}^*$. To account for missing data, we use weighted estimators of the Weibull score equations. We weight by $W_i = (1 - M_i)/(1 - \pi(D_i, \mathbf{X}_i, T_i, Z_i; \boldsymbol{\alpha}^*, \boldsymbol{\tau}))$.

The weighted likelihood for the data is hence

$$\prod_{i=1}^n \{ \exp\{j(\mathbf{X}_i, Z_i; \boldsymbol{\beta}^*)\} \lambda^* \gamma^* t_i^{\gamma^*-1} \}^{W_i D_i} \exp(-W_i \exp\{j(\mathbf{X}_i, Z_i; \boldsymbol{\beta}^*)\} \lambda^* t_i^{\gamma^*})$$

4.1 Relationship of missingness with grade

As τ_1 and τ_2 are increased, the modeled difference in the log odds of grade being missing for worse versus better differentiated grades is presumed to increase, all other covariates being held equal. As τ_1 and τ_2 increase, the proportion of tumors in the missing data assumed to be worse grades also increases. To see this, note that the missingness model implies that for conditional probabilities of missingness when $X_{2i} = 3$ versus $X_{2i} = 1$,

$$\frac{\text{odds } P[M_i=1|D_i, \mathbf{X}_{1i}, X_{2i}=3, T_i, Z_i]}{\text{odds } P[M_i=1|D_i, \mathbf{X}_{1i}, X_{2i}=1, T_i, Z_i]} = \exp(2\tau_1 + 2\tau_2 Z_i)$$

This likewise implies that,

$$\frac{P[X_{2i}=3|M_i=1,D_i,\mathbf{X}_{1i},T_i,Z_i]}{P[X_{2i}=1|M_i=1,D_i,\mathbf{X}_{1i},T_i,Z_i]} / \frac{P[X_{2i}=3|M_i=0,D_i,\mathbf{X}_{1i},T_i,Z_i]}{P[X_{2i}=1|M_i=0,D_i,\mathbf{X}_{1i},T_i,Z_i]} = \exp(2\tau_1 + 2\tau_2 Z_i)$$

and,

$$= \frac{\frac{P[X_{2i}=3|M_i=1,D_i,\mathbf{X}_{1i},T_i,Z_i]}{P[X_{2i}=1|M_i=1,D_i,\mathbf{X}_{1i},T_i,Z_i]}}{\frac{P[X_{2i}=3|M_i=0,D_i,\mathbf{X}_{1i},T_i,Z_i]}{P[X_{2i}=1|M_i=0,D_i,\mathbf{X}_{1i},T_i,Z_i]}} \exp(2\tau_1 + 2\tau_2 Z_i)$$

As τ_1 and τ_2 increase, we would expect that the ratio of grade 3 versus grade 1 tumors in the missing data would increase such that for extreme values of τ , many more tumors are grade 3 than grade 1 tumors. For similar comparisons of conditional probabilities of missingness when $X_{2i} = 2$ versus $X_{2i} = 1$ and $X_{2i} = 3$ versus $X_{2i} = 2$, we would see that we would have a general worsening of grade in the missing data as the sensitivity parameters are increased.

5 Estimation

In describing the estimation of the parameters, we first need to establish notation for the parameters and parameter space of interest. Let the true parameters be

$$\psi^* = (\alpha^*, \beta^*, \gamma^*, \lambda^*)'$$

We assume that $\psi^* \in \Psi$, where Ψ is a compact parameter space containing candidate values of α^* (noted as α), β^* (noted as β), γ^* (noted as γ), λ^* (noted as λ). We use estimating equations to estimate the parameters. The estimator of Ψ , denoted by $\hat{\psi} = (\hat{\alpha}', \hat{\beta}', \hat{\gamma}', \hat{\lambda}')$, can be found by solving the following unbiased estimating equation

$$\sum_{i=1}^n U(O_i; \psi) = 0$$

where $i = 1, \dots, n$ indexes the study participants and

$$\begin{aligned} U(O_i; \psi) = & [\mathbf{U}_\alpha(O_i; \psi)', \mathbf{U}_\beta(O_i; \psi)', \mathbf{U}_\gamma(O_i; \psi)', \mathbf{U}_\lambda(O_i; \psi)']' \\ \mathbf{U}_\alpha(O_i; \psi) = & \frac{\partial h(D_i, \mathbf{X}_{1i}, T_i, Z_i; \alpha)}{\partial \alpha} \left(M_i - \frac{(1-M_i)\pi(D_i, \mathbf{X}_{1i}, T_i, Z_i; \alpha, \tau)}{(1-\pi(D_i, \mathbf{X}_{1i}, T_i, Z_i; \alpha, \tau))} \right) \\ \mathbf{U}_\beta(O_i; \psi) = & \frac{(1-M_i)}{(1-\pi(D_i, \mathbf{X}_{1i}, T_i, Z_i; \alpha, \tau))} \times \\ & \left(\frac{\partial}{\partial \beta} j(\mathbf{X}_i, Z_i; \beta) D_i - \frac{\partial}{\partial \beta} j(\mathbf{X}_i, Z_i; \beta) \exp\{j(\mathbf{X}_i, Z_i; \beta)\} \lambda T_i^\gamma \right) \\ \mathbf{U}_\gamma(O_i; \psi) = & \frac{(1-M_i)}{(1-\pi(D_i, \mathbf{X}_{1i}, T_i, Z_i; \alpha, \tau))} \times \\ & \left(\frac{D_i}{\gamma} + D_i \log T_i - \lambda \exp\{j(\mathbf{X}_i, Z_i; \beta)\} T_i^\gamma \log T_i \right) \\ \mathbf{U}_\lambda(O_i; \psi) = & \frac{(1-M_i)}{(1-\pi(D_i, \mathbf{X}_{1i}, T_i, Z_i; \alpha, \tau))} \left(\frac{D_i}{\lambda} - \exp\{j(\mathbf{X}_i, Z_i; \beta)\} T_i^\gamma \right) \end{aligned}$$

Proofs of identifiability and unbiasedness of the estimating equations are akin to the proofs of Rotnitzky *et al.* [18].

We do not propose the use of a doubly robust estimator in this work as was done by Wang and Chen in the context of Cox models [21] since doing so requires the specification of the conditional distribution of X_{2i} given observed covariates. While doubly robust estimators can be consistent if either the missingness or survival model is correct, the approach is not a panacea as there can potentially be loss of efficiency if the conditional distribution of the missing data given the covariates is incorrect [22]. More importantly to this work, the exact specification of one model when using a sensitivity analysis with respect to another model might seem counterintuitive to many, particularly when a gain in efficiency is not guaranteed.

Solving the estimating equations is a relatively straight forward process. We can first estimate α^* using a Newton-Raphson algorithm for fixed τ . Once we have $\hat{\alpha}$ we can then estimate the probabilities of having missing data. Next, we can obtain estimates of β^* , γ^* , and λ^* using the survey weighted estimator for a Weibull proportional hazards regression found in many statistical software packages. We can do this for a range of τ as part of a sensitivity analysis.

To construct standard errors, we use the commonly termed sandwich variance estimator described by Huber [23]. Hence, the variance of $\hat{\psi}$ can be approximated by an empirical

estimate of $E_{\psi^*} \left[\frac{\partial \mathbf{U}_{(O;\psi)}}{\partial \psi} \right]^{-1} E_{\psi^*} [\mathbf{U}(O;\psi) \mathbf{U}(O;\psi)'] E_{\psi^*} \left[\frac{\partial \mathbf{U}_{(O;\psi)}}{\partial \psi} \right]^{-1}$. By stacking the estimating equations, our robust standard errors for $\hat{\beta}$ account for the estimation of α in the weights used to estimate the Weibull model. The standard errors of $\hat{\beta}$ are hence different from the robust standard errors generated by typical statistical software that does not account for the fact that weights are estimated.

6 Simulation Study

We performed a simulation study to examine the bias that might occur when an investigator fails to account for missing data using large population-based databases such as SEER. We generated data under the following algorithm.

1. Generate a random variable X from a bernoulli distribution with probability 0.60.
2. Generate a random treatment indicator as bernoulli with conditional probability, $P(Z = 1|X) = \text{expit}(-1 + 2X)$.
3. Generate a survival time, Y , in months as exponential with a density of $f(y|X, Z) = \lambda(X, Z) \exp\{-\lambda(X, Z)y\}$ with $\lambda(X, Z) = \exp(-5 + 2Z + X + XZ)$. The log hazard ratio for the effect of the treatment ($Z = 1$) on survival when $X = 0$ is hence 2, and the log hazard ratio for the difference in the effect of treatment on survival when $X = 1$ compared to $X = 0$ is hence 1.
4. Generate a censoring time, C , as exponential with $\lambda = 0.05$. Let $T = \min(Y, C)$ be the observed time to censoring or death and let $D = 1$ if $Y \leq C$, 0 otherwise.
5. Generate a missing data indicator, M , to represent that X is missing as Bernoulli with $P(M = 1|D, X, T, Z) = \text{expit}(-1.5 + Z(T/12)^{0.5} + .5(1 - D)(T/12)^{0.5} + 0.5Z + 0.5X + 2XZ)$. $M = 1$ indicates that X is missing, 0 otherwise. This gives us a $\tau_1^* = .5$ and $\tau_2^* = 2$ for the treatment main effect and interaction terms, respectively.

We generated simulated data samples of 1500. After generating the data, we estimated the missing data and Weibull proportional hazards regression parameters using the sensitivity analysis as described above. We repeated the step 5000 times for values of the sensitivity

parameters that were within a range of 2 of the true parameters. We examined the proportion of 95% confidence intervals that contained the truth.

The algorithm described above results in approximately 55% of the sample being assigned to treatment ($Z = 1$), 48% having a censoring time less than the survival time, and 57% of the sample having missing data. The average time until death is approximately 4.6 years, but the average differs according to treatment, covariate, and missingness patterns. For example, for those with a missing X , average time to death is approximately 2.6 years. However, for those without missing data, average survival is approximately 7.3 years.

Table 1 presents the results of the simulations. We estimated the bias and coverage at the truth and within 2 points above and below the truth. We also estimated the bias and coverage when the sensitivity parameters are set to 0, which would be representative of a naive analysis in which the missing data mechanism is assumed to be missing at random. In the table, we see that the 95% confidence intervals are adequate (94% empirical coverage) and bias is slight when the sensitivity parameters (τ_1 and τ_2) are set to the truth (τ_1^* and τ_2^*). However, the estimates become more biased and the 95% confidence interval coverage worsens as the parameters deviate from the truth. Importantly, when τ_1 and τ_2 are set to zero, the 95% coverage is unacceptably low (68.0% for the main effect term and 87.5% for the interaction term).

The results indicate the importance of the proposed sensitivity analysis. Inferences can be dramatically different when the data are incorrectly assumed to be missing at random. When the sensitivity analysis contains the truth, there is little bias and the 95% confidence intervals give good empirical coverage.

7 SEER Example

We examined the relationship of surgery with survival among the 7,086 individuals 30 years or older with metastatic kidney cancer in the SEER dataset diagnosed between 1988 and 2002 [24]. We excluded 2,836 due to missing tumor size, surgery, or mortality information. The majority (2,562) were excluded due to missing tumor size information. Our final cohort for analysis was 4,250 individuals. Table 2 displays characteristics of the sample stratified by surgical and missing grade status.

Patients who underwent partial, complete, or radical nephrectomy or nephrectomy NOS were classified as undergoing surgery. Patients who underwent only biopsy, exploratory surgery, palliative bypass, or had unknown status were classified as having not received surgery. There were 1,837 (43%) individuals who did not have surgery and 2,413 (57%) who did have surgery. A total of 2,875 (68%) individuals had missing grade data, with 83% of those who did not have surgery missing grade information compared with 52% of those who did have surgery. Grade was defined as specified in the SEER database except for the collapsing of the two highest grades: well differentiated (grade 1), moderately differentiated (grade 2), and poorly differentiated, undifferentiated, or anaplastic (which we label as grade 3).

Demographic information is listed in Table 2. After stratifying by missing grade, those who had surgery were younger on average and had larger tumors. In addition, in the stratified data, the surgery group had fewer women, more married individuals, and fewer non-whites than the non-surgery group. For these analyses, we grouped a small number of individuals (fewer than 25) with missing race information in the non-white category, and 93 individuals with missing marital status into the non-married category. Those with missing grade information were more likely to have been diagnosed at earlier time periods (e.g. pre-1992)

than those with complete grade information. Among those who had grade data available, grade seemed to be worse in those who had surgery than those who did not. Although grade appears to be worse in patients who underwent surgery than those who did not, this may be due to the fact that patients who underwent surgery had more adequate pathologic specimens, allowing for assessment of tumor grade.

In Figure 1 we present Kaplan-Meier curves of mortality outcomes by stage across the 3 grades. We see that those who did not have surgery in the sample had substantially worse outcomes on average than those who did have surgery. This may be due in large part to selection bias, since less healthy patients will likely be considered poor surgical candidates.

In order to control for baseline differences between the groups, we examined the effect of surgery in a Weibull proportional hazards regression using the complete data as presented in Table 3. We included age, sex, tumor size, race, marital status, and year of diagnosis as covariates in the model in addition to surgery, grade, and their interaction. We used restricted cubic spline basis functions [25] to account for the terms for age (2 interior knots), size (2 interior knots), and year of diagnosis (1 interior knot). We entered the covariates similarly into the missing data model for the weighted analysis. Restricted cubic splines allow for nonlinear effects in models. In Table 3, the hazard ratio of the effect of surgery on mortality among those with grade 1 tumors (the main effect term of surgery) is 0.31 (95% CI 0.19-0.52). The hazard ratio increases with increasing grade. In the regression, the interaction terms of surgery with grade are not significant suggesting that grade does not moderate the protective effect of surgery on outcomes. None of the coefficients of the potentially confounding variables was significantly associated with mortality in the model. In the case of the continuous variables, however, the magnitude and interpretability of the effects is dependent on the scale of the spline basis functions.

Fitting the naive regression is similar to assuming that τ_1 and τ_2 equal zero. In Figure 2, we present the hazard ratio over various assumptions about the relationship of surgery and its interaction with grade in the missing data model. For completeness of presentation, we report analyses over a range of sensitivity analysis parameters that includes extreme assumptions (values of τ_1 and τ_2 that when exponentiated include odds ratios of 1/15 to 15). Later in this Section, we will propose a region in which we believe the truth likely to exist.

In the sensitivity analyses, we entered covariates into the missing data model as described in the Weibull model, but also included censoring status, time until death or censoring (using a restricted cubic spline with 1 interior knot), and interactions between censoring status and all covariates and an interaction between surgical status and time until death or censoring. In this way, we sought to create a flexible missing data model.

When $\tau_2 = 0$, we have the no interaction model in which the relationship of grade with missingness does not vary by surgical status. In the no interaction case, when $\tau_1 = 0$, the hazard ratio is approximately 0.3 to 0.35 regardless of grade, which is consistent with the naive Weibull regression presented in Table 3. When τ_1 is negative, the effect of surgery becomes more protective for those with grade 2 tumors. In the no-interaction case, a negative value of τ_1 means that the log odds of having missing data decreases as grade increases. At extreme negative values of τ_1 , this would indicate that much of the missing data consists of better differentiated tumors, as discussed in Section 4.1. Intuitively, incorrect negative specification of τ_1 would result in the incorrect classification of many grade 3 tumors as grade 1 or grade 2 tumors. Since there are fewer grade 1 and grade 2 tumors with observed data, this could explain why varying τ_1 has a larger impact on grade 1 and grade 2 estimates than grade 3 estimates. As τ_1 increases but τ_2 remains equal to zero, more of the missing tumors are assumed to be higher grades. At extreme positive values of

τ_1 , few of the missing tumors are assumed to be grade 1 tumors, as discussed in Section 4.1, and hence the surgical effect among grade 1 tumors would be less affected by the sensitivity analysis. This indeed seems to be the case in Figure 2 regardless of τ_2 .

As τ_2 decreases, missing grades in those with surgery are assumed to be better grades relative to those who did not have surgery. This could affect estimates of the surgical effect as the surgical benefit would be due in part to confounding by grade rather than a true surgical effect. Such an effect could explain why the benefit of surgery decreases dramatically, to the point of non-statistical significance and a hazard ratio greater than 0.60, in those with grade 1 tumors when τ_2 is at extreme negative values. Such confounding could also explain the strengthening of the association when τ_2 is positive but τ_1 is negative.

Overall, the magnitude of the protective surgical effect is generally consistent across the three tumor grades regardless of the sensitivity parameter. The magnitude of the effect generally ranges between 0.3 and 0.5 except in grade 1 tumors when τ_2 is extremely negative and τ_1 is positive.

In terms of statistical significance, the majority of the estimates of the hazard ratio effect are statistically significant. For those with grade 2 and 3 tumors, the p-value is less than 0.01 in all cases. This is likely due to the larger number of individuals with grade 2 and 3 tumors in the completely observed data which reduces the standard errors of the estimates. For those with grade 1 tumors, the significance holds in all cases except for extreme negative values of τ_2 and slightly positive values of τ_1 in which the p-value is above 0.10. Again, the lack of statistical significance could be due to confounding of differential classification of missing grade data between surgical groups as discussed above.

In Figure 3, we present figures representing the main effect and interaction terms from the Weibull proportional hazards used to generate Figure 2. In Figure 3, while there is some evidence of an interaction between surgery and grade in the region in which the surgical effect loses statistical significance in those with grade 1 disease, the p-values for the interaction terms do not fall below 0.05.

Of note is that we do not know the true values of τ_1 and τ_2 . However, it is possible to postulate reasonable bounds for τ_1 and τ_2 . We believe that in this population of patients with metastatic disease, those with worse grade (grade 3) tumors are more likely to be sicker on presentation. Therefore, they may be less likely to undergo aggressive procedures such as cytoreductive nephrectomy and have sufficient tissue to properly identify grade. This suggests that τ_1 would be positive ($\tau_1 > 0$). However, it is likely that such a trend is not as pronounced among those who have undergone surgery as an adequate surgical specimen would reduce the association between grade and missing data. This would result in τ_2 attenuating the relationship of grade with missing data ($-\tau_1 < \tau_2 < 0$). This would suggest that the true values of τ_1 and τ_2 reside in a triangle in the lower right quadrants of the graphs in Figures 2 and 3. The location of the truth in such a region would suggest that surgery is associated with better mortality outcomes ($p < 0.05$ in all cases). However, the absolute magnitude of the protective effects would not be as great as implied by the naive analysis.

8 Discussion

Missing data can be a significant problem when conducting research with observational or administrative data. In those with metastatic kidney disease in the SEER registry database, over half had missing tumor grade information. The large amount of missing data might lead many to question the validity of inferences drawn from using the complete data or accounting for missing data using imputation techniques that assume the data are missing at

random. This paper demonstrates that the method of sensitivity analysis can be used to demonstrate how data that are missing not at random might be affecting inferences.

In our data example, when using the complete case data, the effect of surgery on survival among those with metastatic kidney cancer was highly protective. Tumor grade did not moderate the impact of surgery on outcomes. In sensitivity analyses, inferences were not highly dependent on the nature of a non-random missing data mechanism. However, there were regions in which the surgical effect was either better or worse than implied by the naive analysis. In particular, the effect of surgery on outcomes was not statistically significant in those with grade 1 disease when worse grade was slightly associated with having missing data among those who did not have surgery but the relationship was attenuated or reversed among those who had surgery. While such a result could be related to the small number of grade 1 tumors observed in the data, such a lack of an effect among well differentiated tumors may be clinically plausible. Well differentiated tumors in general may be less aggressive and associated with better outcomes, and hence their removal may have less of an impact on those with metastatic disease than the removal of moderately or poorly differentiated tumors. However, the area of nonsignificance was just outside of a plausible region for the true sensitivity parameters to reside.

The net evidence suggests that surgery is indeed protective of mortality in those with moderately or poorly differentiated tumors. For researchers interested in the impact of grade on outcomes, resources would best be allocated to improving data collection among those with well differentiated tumors to estimate the effect in this group with more confidence. It may be clinically useful to know if some patients might be able to forego surgery because of a demonstrable lack of effectiveness. Still, the extreme assumptions necessary for inferences to change in those with well differentiated tumors may not be plausible to some, and hence the data presented might reassure the reader of the utility of surgery in this group.

There are limitations to this research. For one, we focused on inferences among those with complete tumor size data. A substantial portion of the sample was also missing size data and did not contribute information to this study. A further sensitivity analysis could investigate how potentially non-random missing tumor size data could impact our inferences. One problem with accounting for multiple missing covariates is that the dimensionality of the sensitivity analysis increases. In our example, we only present inferences over two sensitivity parameters. The number of parameters could increase to six if we also included sensitivity parameters for tumor size, its interactions with surgery and grade, and the three-way interaction between surgery, tumor size, and grade.

It is important to note that we could not control for all of the potential confounders between surgical assignment and outcomes, such as comorbidities, which are not available in the SEER database. Hence, our analysis cannot be considered a fully causal analysis. However, the ease of using and accessing SEER data to assess broad trends and generate hypotheses about population-level cancer outcomes makes it a useful tool in cancer research, and we believe that it is important that investigators understand the potential impact of missing data on their results.

Our simulation study demonstrates that our method for accounting for missing data when conducting survival analyses gives appropriate empirical confidence interval coverage and consistency when the sensitivity parameters are set to the truth, but also shows that inferences can be biased when the sensitivity parameters are incorrectly specified.

For investigators conducting research on observational or administrative data with large amounts of missing data, we believe that sensitivity analyses of the type described in this

paper should become standard. Such sensitivity analyses would ensure that inferences are not biased by missing data.

Acknowledgments

We thank Sam Litwin, Hua Min, and the reviewers for their helpful comments. Research was supported in part by NIH grant P30 CA 06927 and an appropriation from the Commonwealth of Pennsylvania.

References

- Berndt SI, Carter HB, Schoenberg MP, Newschaffer CJ. Disparities in treatment and outcome for renal cell carcinoma among older black and white patients. *Journal of Clinical Oncology*. 2007; 25(24):3589–3595. [PubMed: 17704406]
- Wong YN, Mitra N, Hudes G, Localio R, Schwartz JS, Wan F, Montagnet C, Armstrong C. Survival associated with treatment versus observation of localized prostate cancer in elderly men. *JAMA*. 2006; 296:2683–2693. [PubMed: 17164454]
- Shahinian VB, Kuo YF, Freeman JL, Goodwin JS. Risk of fracture after androgen deprivation for prostate cancer. *New England Journal of Medicine*. 2005; 352:154–164. [PubMed: 15647578]
- Herring AH, Ibrahim JG, Lipsitz SR. Non-ignorable missing covariate data in survival analysis: A case-study of an International Breast Cancer Study Group trial. *Applied Statistics*. 2004; 53(2):293–310.
- Flanigan RC, Salmon SE, Blumenstein BA, Bearman SI, Roy V, McGrath PC, Caton JR Jr, Munshi N, Crawford ED. Nephrectomy Followed by Interferon Alfa-2b Compared with Interferon Alfa-2b Alone for Metastatic Renal-Cell Cancer. *New England Journal of Medicine*. 2001; 345:1655–1659. [PubMed: 11759643]
- Motzer RJ, Mazumdar M, Bacik J, Berg W, Amsterdam A, Ferrara J. Survival and Prognostic Stratification of 670 Patients With Advanced Renal Cell Carcinoma. *Journal of Clinical Oncology*. 1999; 17:2530. [PubMed: 10561319]
- Escudier B, Eisen T, Stadler WM, Szczylik C, Oudard S, Siebels M, Negrier S, Chevreau C, Solska E, Desai AA, Rolland F, Demkow T, Hutson TE, Gore M, Freeman S, Schwartz B, Shan M, Simantov R, Bukowski RM. Sorafenib in Advanced Clear-Cell Renal-Cell Carcinoma. *New England Journal of Medicine*. 2007; 356:125–134. [PubMed: 17215530]
- Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Rixe O, Oudard S, Negrier S, Szczylik C, Kim ST, Chen I, Bycott PW, Baum CM, Figlin RA. Sunitinib versus Interferon Alfa in Metastatic Renal-Cell Carcinoma. *New England Journal of Medicine*. 2007; 356:115–124. [PubMed: 17215529]
- Pantuck AJ, Belldegrun AS, Figlin RA. Cytoreductive Nephrectomy for Metastatic Renal Cell Carcinoma: Is It Still Imperative in the Era of Targeted Therapy? *Clinical Cancer Research*. 2007; 13:693s–696. [PubMed: 17255295]
- Halbert RJ, Figlin RA, Atkins MB, Bernal M, Hutson TE, Uzzo RG, Bukowski RM, Khan KD, Wood CG, Dubois RW. Treatment of patients with metastatic renal cell cancer. *Cancer*. 2006; 107:2375–2383. [PubMed: 17048248]
- Hudes G.; Carducci, M.; Tomczak, P.; Dutcher, J.; Figlin, R.; Kapoor, A.; Staroslawska, E.; Sosman, J.; McDermott, D.; Bodrogi, I.; Kovacevic, Z.; Lesovoy, V.; Schmidt-Wolf, IG.; Barbarash, O.; Gokmen, E.; O'Toole, T.; Lustgarten, S.; Moore, L.; Motzer, RJ. A phase 3, randomized, 3-arm study of temsirolimus (TEMSR) or interferon-alpha (IFN) or the combination of TEMSR + IFN in the treatment of first-line, poor-risk patients with advanced renal cell carcinoma (adv RCC). *Journal of Clinical Oncology*, 2006 ASCO Annual Meeting Proceedings Part I; 2006. p. LBA4
- Rioux-Leclercq N, Rioux-Leclercq N, Karakiewicz PI, Trinh QD, Ficarra V, Cindolo L, de la Taille A, Tostain J, Zigeuner R, Mejean A, Patard JJ. Prognostic ability of simplified nuclear grading of renal cell carcinoma. *Cancer*. 2007; 109(5):868–874. [PubMed: 17262800]
- Patard JJ, Kim HL, Lam JS, Dorey FJ, Pantuck AJ, Zisman A, Ficarra V, Han KR, Cindolo L, De La Taille A, Tostain J, Artibani W, Dinney CP, Wood CG, Swanson DA, Abbou CC, Lobel B, Mulders PF, Chopin DK, Figlin RA, Belldegrun AS. Use of the University of California Los

- Angeles Integrated Staging System to Predict Survival in Renal Cell Carcinoma: An International Multicenter Study. *Journal of Clinical Oncology*. 2004; 22:3316–3322. [PubMed: 15310775]
14. Rubin DB. Inference and missing data. *Biometrika*. 1976; 63:581–592.
 15. Schafer, JL. Analysis of incomplete multivariate data. Boca Raton, FL: Chapman and Hall; 1997.
 16. Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*. 1995; 90:106–121.
 17. Rathouz PJ. Identifiability assumptions for missing covariate data in failure time regression models. *Biostatistics*. 2007; 8(2):345–356. [PubMed: 16840561]
 18. Rotnitzky A, Robins J, Scharfstein D. Semiparametric regression for repeated outcomes with non-ignorable non-response. *Journal of the American Statistical Association*. 1998; 93:1321–1339.
 19. Scharfstein D, Rotnitzky A, Robins J. Adjusting for non-ignorable drop-out using semiparametric drop-out models. *Journal of the American Statistical Association*. 1999; 94:1096–1120.
 20. Baker SG, Ko CW, Graubard BI. A sensitivity analysis for nonrandomly missing categorical data arising from a national health disability survey. *Biostatistics*. 2003; 4(1):41–56. [PubMed: 12925329]
 21. Wang CY, Chen HY. Augmented Inverse Probability Weighted Estimator for Cox Missing Covariate Regression. *Biometrics*. 2001; 57(2):414–419. [PubMed: 11414564]
 22. Carpenter JR, Kenward MG, Vansteelandt S. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society, Series A*. 2006; 169(3):571584.
 23. Huber PJ. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*. 1964; 35(1):73–101.
 24. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 9 Regs Limited-Use, Nov 2004 Sub (1973-2002), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch released April 2005, based on the November 2004 submission.
 25. Harrell, FE. Regression Modeling Strategies. Vol. Chapter 2. New York, NY: Springer; 2001.

Figure 1.
Kaplan-Meier Survival Estimates by Grade in the Observed Data

Figure 2. Hazard ratio (contour lines) of surgical effect in the Weibull proportional hazards regressions for various values of the sensitivity parameters, τ_1 (the main effect term) and τ_2 (the interaction term).

Figure 3. Hazard ratios (contour lines) representing the main effect and interaction terms in the Weibull proportional hazards regressions for various values of the sensitivity parameters, τ_1 (the main effect term) and τ_2 (the interaction term).

Table 1

Results of simulation study

τ_1	τ_2	Main Effect Haz. Ratio	Coverage of 95% C.I.	Interaction Haz. Ratio	Coverage of 95% C.I.
0.5 (truth)	2.0 (truth)	2.04	0.941	1.02	0.940
0.0	0.0	2.33	0.680	0.77	0.875
0.0	2.0	2.08	0.932	0.95	0.935
0.0	4.0	2.12	0.903	0.85	0.876
0.5	0.0	2.25	0.813	0.87	0.927
0.5	4.0	2.15	0.872	0.88	0.886
-1.5	0.0	2.48	0.341	0.60	0.709
-1.5	2.0	2.25	0.800	0.80	0.886
-1.5	4.0	2.08	0.930	0.85	0.905
2.5	0.0	1.89	0.946	1.36	0.837
2.5	2.0	1.97	0.944	1.28	0.870

Table 2

Characteristics of sample.

	Missing Grade		Grade Available	
	Surgery	No Surgery	Surgery	No Surgery
N	872	2003	965	410
Age (SD)	62.0 (11.5)	68.0 (12.5)	61.6 (11.4)	66.0 (12.1)
Size in cm (SD)	8.5 (4.3)	8.4 (4.0)	9.1 (4.6)	8.1 (4.2)
Female	29%	39%	32%	35%
Married	71%	57%	69%	63%
Non-white	12%	17%	13%	15%
Diagnosed 1988-1992	27%	39%	22%	22%
Diagnosed 1993-1997	36%	27%	41%	27%
Diagnosed 1998-2002	38%	34%	37%	51%
Grade=1			5%	10%
Grade=2			28%	24%
Grade=3 or grade=4			66%	65%

Table 3

Hazard ratios from Weibull model fit on those with complete data (n=1,375)

Variable	Hazard Ratio	95% CI	p-value
Age	0.99	(0.97, 1.01)	0.339
Age spline term 1	1.05	(0.99, 1.11)	0.131
Age spline term 2	0.86	(0.68, 1.09)	0.202
Size	1.01	(1.00, 1.01)	0.151
Size spline term 1	0.99	(0.97, 1.02)	0.541
Size spline term 2	1.02	(0.95, 1.10)	0.600
Year diagnosed	1.00	(0.95, 1.05)	0.897
Year spline term 1	1.00	(0.95, 1.05)	0.981
Female	1.07	(0.93, 1.25)	0.347
Married	1.08	(0.92, 1.27)	0.331
Non-white	0.96	(0.78, 1.17)	0.665
Surgery	0.31	(0.19, 0.52)	<0.001
Grade 2	1.46	(0.93, 2.28)	0.101
Grade 3	2.55	(1.71, 3.81)	<0.001
Surgery * Grade 2	1.04	(0.58, 1.88)	0.886
Surgery * Grade 3	1.15	(0.67, 2.00)	0.610
γ	0.90	(0.87, 0.94)	
λ	0.57	(0.30, 1.08)	