

# *University of North Carolina at Chapel Hill*

The University of North Carolina at Chapel Hill Department of  
Biostatistics Technical Report Series

---

*Year 2013*

*Paper 34*

---

## Parameter Estimation in Cox Proportional Hazard Models with Missing Censoring Indicators

Naomi Brownstein\*

Eric Bair†

Jianwen Cai‡

Gary Slade\*\*

\*University of North Carolina at Chapel Hill, nbrownst@live.unc.edu

†University of North Carolina at Chapel Hill, ebair@email.unc.edu

‡University of North Carolina at Chapel Hill, cai@bios.unc.edu

\*\*University of North Carolina at Chapel Hill, gary\_slade@dentistry.unc.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/uncbiostat/art34>

Copyright ©2013 by the authors.

# Parameter Estimation in Cox Proportional Hazard Models with Missing Censoring Indicators

Naomi Brownstein, Eric Bair, Jianwen Cai, and Gary Slade

## Abstract

In a prospective cohort study, examining all participants for incidence of the condition of interest may be prohibitively expensive. For example, the “gold standard” for diagnosing temporomandibular disorder (TMD) is a clinical examination by an expert dentist. In a large study, examining all subjects in this manner is infeasible. Instead, it is common to use a cheaper (and less reliable) examination to screen for possible incident cases and perform the “gold standard” examination only on participants who screen positive on this simpler examination. Unfortunately, some subjects may leave the study before receiving the “gold standard” examination. Within the framework of survival analysis, this results in missing censoring indicators. Motivated by the Orofacial Pain: Prospective Evaluation and Risk Assessment (OPPERA) study, a large cohort study of TMD, we propose a method for parameter estimation in survival models with missing censoring indicators. We estimate the probability of being a case for those with no “gold standard” examination using logistic regression. These predicted probabilities are used to generate multiple imputations of each missing case status and estimate the hazard ratios associated with each putative risk factor. The variance introduced by the procedure is estimated using multiple imputation. We simulate data with missing censoring indicators and show that our method performs as well as or better than the competing methods. Finally, we apply the proposed method to analyze data from the OPPERA study.

## Parameter Estimation in Cox Proportional Hazard Models with Missing Censoring Indicators

Naomi Brownstein<sup>1,\*</sup>, Eric Bair<sup>1,2,\*\*</sup>, Jianwen Cai<sup>1,\*\*\*</sup>, Gary Slade<sup>2,\*\*\*\*</sup>

<sup>1</sup>Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill NC, U.S.A.

<sup>2</sup>School of Dentistry, University of North Carolina at Chapel Hill, Chapel Hill NC, U.S.A.

\**email*: nbrownst@mail.bios.unc.edu

\*\**email*: ebair@email.unc.edu

\*\*\**email*: cai@bios.unc.edu

\*\*\*\**email*: gary\_slade@dentistry.unc.edu

**SUMMARY:** In a prospective cohort study, examining all participants for incidence of the condition of interest may be prohibitively expensive. For example, the “gold standard” for diagnosing temporomandibular disorder (TMD) is a clinical examination by an expert dentist. In a large study, examining all subjects in this manner is infeasible. Instead, it is common to use a cheaper (and less reliable) examination to screen for possible incident cases and perform the “gold standard” examination only on participants who screen positive on this simpler examination. Unfortunately, some subjects may leave the study before receiving the “gold standard” examination. Within the framework of survival analysis, this results in missing censoring indicators. Motivated by the Orofacial Pain: Prospective Evaluation and Risk Assessment (OPPERA) study, a large cohort study of TMD, we propose a method for parameter estimation in survival models with missing censoring indicators. We estimate the probability of being a case for those with no “gold standard” examination using logistic regression. These predicted probabilities are used to generate multiple imputations of each missing case status and estimate the hazard ratios associated with each putative risk factor. The variance introduced by the procedure is estimated using multiple imputation. We simulate data with missing censoring indicators and show that our method performs as well as or better than the competing methods. Finally, we apply the proposed method to analyze data from the OPPERA study.

**KEY WORDS:** Cox regression; missing data; multiple imputation; survival analysis.

## 1. Introduction

Time-to-event analyses are frequently conducted in medicine, actuarial science, and numerous other fields of applied science. There is a well-developed set of survival analysis methods implemented in standard software. Semi-parametric methods, such as the Cox proportional hazards model, allow robust estimation of the effects of covariates on the hazard function. Yet, these methods require the analyst to know not only the time of the event or censoring, but also the event status, which may not always be available.

Often, the outcome of interest may be difficult to ascertain. For example, in oncology studies, researchers may want to differentiate between deaths due to cancer and deaths due to car accidents or other unrelated causes. Investigators may easily record the mortality of all subjects, but it may be extremely difficult or costly to find out exactly why each subject died. One possible solution to this problem is delayed event adjudication (Cook and Kosorok, 2004). This means that possible cases are not identified immediately but screened using simple methods that may have poor sensitivity or specificity; then, one or more experts examines the screened candidate cases using a more precise, but also more costly and time-consuming method to determine the true event status.

When individuals have more than one potential event, Snapinn (1998) recommends screening out unlikely events and using weights for each potential event and use a weighted partial likelihood. Each observation is weighted by the probability that the subject had first failed at the observed time; these probabilities may be known or estimated. Taking this uncertainty into account reduces bias and increases power. However, estimating the weights depends upon the presence of a normally distributed diagnostic variable and either knowing or having experts guess the relative frequency of true endpoints to false endpoints. In addition, the variance of the method is mentioned but not described in detail.

The study that motivates our work is Orofacial Pain: Prospective Evaluation and Risk

Assessment (OPPERA), a prospective cohort study to identify risk factors for the onset of temporomandibular disorders (TMD). A total of 3,263 initially TMD-free subjects were recruited at four studies sites between 2006 and 2008. A specialized dental examination (Dworkin and LeResche, 1992) is required for accurately diagnosing TMD. It is impractical to perform this exam on every participant in such a large study, especially if most study participants are unlikely to have the condition. Instead, the “gold standard” examination is performed only on subjects who screen positive on a simpler exam, such as a survey designed to assess recent orofacial pain (Slade et al., 2012). However, some subjects who screen positive may be lost to follow-up before receiving the “gold standard” examination. Subjects who screen positive but fail to return for their follow-up examination may differ from those who do complete the examination according to known demographic factors. A time-to-event analysis would then have some subjects with missing censoring indicators, which are missing at random but not missing completely at random. This setting presents statistical challenges, which require care in order to avoid bias and maintain efficiency.

Over the past three decades, a variety of estimators of the survival function for time-to-event data with missing censoring indicators have been proposed. Dinse (1982), Racine-Poon and Hoel (1984), and Flehinger et al. (1998) propose estimates for competing risk models. van der Laan and McKeague (1998) first provide an asymptotically efficient nonparametric estimator of the survival function for standard univariate failure time data with censoring indicators missing at random. Subramanian and Bean (2008b), Subramanian (2009), and Subramanian (2011) discuss improving the survival function estimate via bootstrapping and multiple imputation. Additional research on estimating the survival function include Subramanian (2004), Subramanian (2006), Subramanian and Bean (2008a), and Subramanian and Bandyopadhyay (2010). In addition, the logrank test has been generalized for missing censoring indicators (Goetghebeur and Ryan, 1990; Dewanji, 1992; Tsiatis et al., 2002).

Less is known about the more difficult problem of fitting a Cox proportional hazard model in the presence of missing censoring indicators. Progress began in the competing risk setting. Goetghebeur and Ryan (1995) propose competing risks proportional hazards regression models using estimating equations assuming that observations are known either to be censored or to have failed by some cause. They assume proportional hazards for each failure type and between the two failure types. For the linear transformation model, Gao and Tsiatis (2005) propose an augmented inverse probability weighted estimator and algorithm and illustrate its asymptotic and double robustness properties.

For the usual survival setting with only one cause of failure and possibly missing censoring indicators, McKeague and Subramanian (1998) and Gijbels et al. (2007) propose estimating functions for the estimation of the parameters in the Cox proportional hazards models. Subramanian (2000) conducts parameter estimation in the Cox model under the assumption of proportional censorship. Yet, these methods depend on the MCAR assumption, which may not be realistic in practice and does not hold for the OPPERA study. Chen et al. (2009) estimate Cox regression parameters using the EM algorithm and establish their consistency under basic regularity conditions, including MAR censoring indicators. However, their approach depends on the assumptions of piecewise constant proportional hazard functions for the censoring time as well as for the failure time.

Cook and Kosorok (2004) estimate parameters in standard Cox proportional hazard models with observations weighted by their probability of being a true case and provide a formula for estimation of the variance of the estimators. They showed that estimators are consistent and asymptotically normally distributed. However, the formula for the variance of their estimator did not produce reasonable results for the OPPERA study. Alternatively, they also propose approximating the standard error using bootstrapping.

Every three months, OPPERA study participants are asked to complete a quarterly health

update (QHU) of recent pain (Slade et al., 2012). Subjects who screen positive are contacted to schedule a follow-up dental examination, based on the Research Diagnostic Criteria (RDC) for temporomandibular disorders (Dworkin and LeResche, 1992). This RDC exam is considered to be the “gold standard” for diagnosing TMD.

The likelihood that a participant will come in for their RDC exam differs based on known factors such as gender, race, or socio-economic status. This indicates that the censoring indicators in the OPPERA study are not missing completely at random. Application of models that assume MCAR censoring indicators may result in biased estimates of hazard ratios for covariates of interest. More significantly, a subject’s responses to their QHU are predictive of whether or not they are a case. There is a clear need for new methodology to effectively answer the research questions of the OPPERA study.

In this paper, we propose a method for parameter and variance estimation in the Cox regression model with missing censoring indicators. We describe our method in section 2. In section 3, we report the results of simulations. Finally, in section 4 we apply our method to the OPPERA study. We conclude with a discussion in section 5.

## 2. Model

Assume there are  $n$  independent subjects. For each subject,  $i = 1, \dots, n$ , let  $C_i$  and  $T_i$  denote the potential times until censoring and failure, respectively,  $V_i = \min(T_i, C_i)$ ,  $\Delta_i = I(T_i \leq C_i)$ ,  $N_i(t) = I(T_i < t)$ ,  $Z_i$  a  $p \times 1$  vector of covariates and  $X_i$  a  $q \times 1$  vector of auxiliary covariates. We assume the hazard for subject  $i$  follows a Cox proportional hazards model

$$\lambda(t|z_i) = \lambda_0(t) \exp(\beta' z_i) \quad (1)$$

where  $\lambda_0(t)$  is an unspecified baseline hazard function. Let  $\xi_i$  denote the indicator that  $\Delta_i$  is observed and  $\sigma_i = \xi_i \Delta_i$ . We observe  $(V_i, \xi_i, \sigma_i)$  for  $i = 1, \dots, n$ .

In the OPPERA study,  $V_i$  is the length of time since subject  $i$  is enrolled in the study until

either he or she screens positive on a QHU and is subsequently diagnosed with TMD or is lost to follow-up. (Note that the study protocol requests that subjects fill out QHUs until the end of the study, even if the subject is diagnosed with TMD. For the purposes of this paper, we do not consider QHUs for subjects who have already been diagnosed with TMD.) If subject  $i$  either screened negative on the last recorded QHU or screened positive on the QHU and had a negative RDC exam, then  $\Delta_i = 0$ . If subject  $i$  screened positive on the QHU and had a positive RDC exam, then  $\Delta_i = 1$ . If subject  $i$  screened positive on the QHU but did not come in for the RDC exam, then  $\Delta_i$  is missing and  $\xi_i = 0$ . If  $\Delta_i$  is observed, then  $\xi_i = 1$ . Measured when a participant enrolled in the study, putative risk factors for TMD are denoted by  $Z_i$ , and a participant's responses to their QHU are denoted by  $X_i$ . For OPPERA, we also define  $Q_i = 1$  if subject  $i$  screens positive on the last QHU and  $Q_i = 0$  otherwise.

We assume the censoring indicators are missing at random (MAR) as follows:

$$\begin{aligned} & P(\xi_i = 0 | V_i, \Delta_i, X_i, Q_i = 1) & (2) \\ = & P(\xi_i = 0 | X_i, V_i, Q_i = 1) = \rho(X_i, V_i) \end{aligned}$$

We model the probability that subject  $i$  with a missing censoring indicator is a case by a logistic regression model based on  $X_i$  and  $V_i$ , i.e.

$$p(\Delta_i = 1 | V_i, X_i, Z_i, \xi_i = 0, Q_i = 1) \quad (3)$$

$$= p(\Delta_i = 1 | X_i, V_i, Q_i = 1) \quad (4)$$

$$= \frac{\exp(\alpha' X_i + \gamma V_i)}{1 + \exp(\alpha' X_i + \gamma V_i)} I(Q_i = 1) \quad (5)$$

That is, we predict the probability that a subject who did not come in for the RDC exam actually has TMD based on the time of the QHU and their answers on the QHU. In the OPPERA study, we observed empirically that the probability of having developed TMD was not associated with baseline covariates once we controlled for the responses on the QHU.



Then the predicted probability of being a case is given by

$$\begin{aligned} p_i &= \hat{E}[\Delta_i | \xi_i, \sigma_i, X_i, Z_i, V_i; \alpha, \beta, \gamma] \\ &= \sigma_i + (1 - \xi_i) \frac{\exp(\hat{\alpha}' X_i + \hat{\gamma} V_i)}{1 + \exp(\hat{\alpha}' X_i + \hat{\gamma} V_i)}. \end{aligned} \quad (6)$$

Note, if there are repeated measures, we may use a generalized linear mixed model rather than a logistic regression model. For example, if a majority of subjects repeatedly screen positive on their QHUs and come in for at least one RDC exam, then we would have multiple observations per subject. In that case, fitting a mixed model rather than a logistic regression model would account for correlations between responses of the same subject.

Once we have the predicted failure probabilities, we may use them to create imputed datasets. We then use the completed data to fit a standard Cox proportional hazards model and estimate the parameters and their standard errors via multiple imputation.

### 2.1 Multiple Imputation

One popular method of parameter and standard error estimation is multiple imputation. Widely available in statistical software, multiple imputation is commonly used by both statisticians and non-statisticians alike. For a comprehensive reference on multiple imputation, see the review paper of Rubin (1996). Our imputation procedure is as follows:

- (i) Estimate predicted probabilities as described in the previous section.
- (ii) Retain each observation with observed censoring indicators.
- (iii) For each observation with a missing censoring indicator, generate a Bernoulli random variable with success probability equal to the predicted probability found in step (i).
- (iv) Combine the raw and imputed data from steps (ii) and (iii) to form a completed dataset.
- (v) Fit the Cox proportional hazards model to the completed dataset.
- (vi) Record each parameter estimate  $\hat{\beta}_j$  and covariance matrix  $\hat{U}_j$ .

(vii) Repeat steps (ii)-(vi)  $m$  times.

Next, we combine all of the estimates. The average parameter estimate is

$$\bar{\beta} = \frac{1}{m} \sum_{j=1}^m \hat{\beta}_j, \quad (7)$$

the within-imputation variance estimate is

$$\bar{U} = \frac{1}{m} \sum_{j=1}^m \hat{U}_j, \quad (8)$$

and the between-imputation variance

$$\hat{B} = \frac{1}{m-1} \sum_{j=1}^m (\hat{\beta}_j - \bar{\beta})(\hat{\beta}_j - \bar{\beta})'. \quad (9)$$

Finally, the estimated covariance matrix is

$$\widehat{Var}(\bar{\beta}) = \bar{U} + \left(1 + \frac{1}{m}\right) \hat{B}. \quad (10)$$

## 2.2 Estimation of Incidence

We can also estimate incidence rates using Poisson regression instead of Cox regression. In order to estimate incidence, we estimate the predicted case probabilities as described previously. Then we impute case status as described in section 2.1 but we fit Poisson regression models, rather than Cox models, to the completed datasets. Finally, we calculate the incidence rate based on the estimates of the regression coefficients in the Poisson model.

## 3. Simulations

We simulated data with missing censoring indicators and compared methods with respect to bias, coverage, and confidence interval width. In correspondence with the formulae in Bender et al. (2005) for generating exponentially distributed failure times under the proportional hazards model with covariates, we generated 1000 survival times. That is, the survival time for each individual was distributed according to equation (1) where  $\lambda_0(t) = 1$  is the baseline hazard. For our simulations, the  $Z_i = X_{i1}$  was a single covariate following a normal distribution with mean 2 and unit variance. In other words, the failure times,  $T_i$  followed

an exponential distribution with hazard  $\exp(\beta'X_{i1})$ . We used  $\beta \in \{-0.5, -1.5, -3\}$ . The censoring times,  $C_i$  followed an exponential distribution with mean 5. This yielded about 35%, 75% and 90% censoring, respectively.

For each observation, we generated a random normal variable,  $X_{i2} \sim N(\Delta_i, 0.3)$ . Larger values of  $X_{i2}$  were more likely to be associated with a failure event. We use  $X_{i2}$ , a continuous measure of the potential of being as a case, to generate the indicator  $Q_i = I(X_{i2} > 0.5)$ .

We created missing censoring indicators under the following classical missing data mechanisms of Rubin (1976):

- (I) The probability of having a missing censoring indicator depends on an observed covariate. This is known as missing at random (MAR).
- (II) The probability of having a missing censoring indicator depends on the censoring indicator. This is known as missing not at random (MNAR).

Observations were potentially missing if  $Q_i = 1$ . For mechanism I, we set censoring indicators to be missing based on  $x_{i1}$ , i.e. they were set to missing with probability

$$\rho_i(X_i, V_i) = P(\xi_i = 0 | X_i, V_i, Q_i = 1) = \frac{\exp(-0.25x_{i1})}{1 + \exp(-0.25x_{i1})}.$$

For mechanism II, we set 30% of the censored observations and 50% of the failures to have missing indicators.

Within the framework of OPPERA, for person  $i$ ,  $V_i = \min(T_i, C_i)$  corresponds to the time of the last positive QHU or loss to follow-up,  $\xi_i = I(\Delta_i \text{ is observed})$  to the indicator of whether person  $i$  came in for their RDC exam if  $Q_i = 1$ ,  $\Delta_i$  to the indicator of whether person  $i$  was diagnosed with TMD,  $X_{i1}$  to a risk factor for TMD measured at baseline,  $X_{i2}$  to a covariate collected on the last QHU, and  $Q_i$  as an indicator of whether the person screened positive on their last QHU.

In OPPERA,  $Q_i = 0$  means subject  $i$  had a negative last QHU. Assuming subject  $i$  was not a matched control, he or she would not be subjected to an RDC exam, and therefore

would not be considered a case of TMD. However, about 5% of matched controls had TMD, i.e. there was a false negative rate of 5%. In order to parallel the set-up of OPPERA, we set  $\Delta_i = 0$  if  $Q_i = 0$ ; this corresponded to a false negative rate of about 5% for these simulations.

We first fit all observations with observed censoring indicators in each simulated dataset to a logistic regression model for case status with the covariates  $X_{i1}$  and  $X_{i2}$ . That is, using the complete data (i.e. observations with  $\xi_i = 1$ ), we fit the logistic regression model for the event probability conditional on  $X_i' = (1, X_{i1}, X_{i2})$ , namely

$$\text{logit}\{Pr(\Delta_i = 1|Q_i = 1, \xi_i = 1)\} = \alpha' X_i \quad (11)$$

We output the predicted probabilities,  $\hat{p}_i = \frac{\hat{\alpha}' X_i}{1 + \hat{\alpha}' X_i}$ ,  $i = 1, \dots, 1000$

To evaluate the performance of our method, we used multiple imputation to create 100 datasets for each simulation as described in Section 2.1. That is, we generated 100 datasets with identical data for  $X_i, T_i, C_i, \xi_i$ . For each observation  $i$  with  $\xi_i = 0$ , we generated failure indicators  $\hat{\Delta}_{ij} \sim \text{Bernoulli}(\hat{p}_i)$  independently for each imputation  $j$ .

We fit a Cox proportional hazards model to each dataset completed by multiple imputation and recorded the multiple imputation estimates of the regression coefficient and its variance. These were aggregated using equations (7) and (10) to create confidence intervals for the multiple imputation estimates.

We compared the performance of our method with the performance of the method of Cook and Kosorok (2004). For each simulated data set, we calculated the predicted probabilities  $\hat{p}_i$  that (unobserved) event  $i$  is an event, as described previously. We then fit a weighted Cox proportional hazards model to the dataset. For observations with missing censoring indicators we created two new observations with the same failure time and covariates, but different failure indicators and weights: the first with  $\hat{\Delta}_i = 1$  and weight  $\hat{p}_i$ , the second with  $\hat{\Delta}_i = 0$  and weight  $\hat{p}_i$ . We used unit weight for subjects with fully observed data and recorded the estimated regression coefficient,  $\hat{\beta}$ . We estimated the variance of this estimate

by generating 1000 bootstrap replicates of each simulated data set. We recorded the average parameter estimate,  $\bar{\hat{\beta}}$  and percentile confidence intervals  $(\beta_{0.025}, \beta_{0.975})$ . Here,  $\beta_\alpha$  is the  $\alpha$  quantile among the 1000 bootstrap replicates.

We also compared our method to the ideal situation in which all data were observed, complete case analysis (meaning that we exclude all observations with missing censoring indicators from the data set), and two ad-hoc methods in which we treat the missing indicators either all as censored or all as failures. Results under the assumption of MAR are included in Table 1. We estimated the bias of each method by calculating the mean difference between the estimated Cox regression coefficient and the true coefficient over the 1000 simulations. We also calculated the mean width of the confidence intervals produced by each method over the 1000 simulations. Similarly, we calculated the empirical coverage probability for the confidence intervals produced by each method by dividing the number of times that the confidence intervals contained the true value of the parameter by 1000. We also tested if the coverage probabilities differed significantly from the nominal rate and if the width of the confidence intervals were different between our method and the comparison methods.

[Table 1 about here.]

The imputed confidence intervals achieved the desired coverage probability in all simulations. Our multiple imputation method and the method of Cook and Kosorok (2004) produced unbiased estimates and valid confidence intervals in all the scenarios we considered. The estimates produced by the other methods showed a small (but statistically significant) amount of bias and did not always achieve the desired amount of coverage. Our multiple imputation method also yielded the narrowest confidence intervals in each scenario, although the method of Cook and Kosorok (2004) produced confidence intervals that were only slightly

wider. Moreover, for most parameter values, the coverage probabilities for the complete case and ad hoc methods were significantly different ( $p < 0.01$ ) from the nominal rate.

In addition, we examined the performance of our proposed methods when we changed the logistic regression model for  $\Delta_i$ . We investigate two additional types of models: one in which the model contained a variable unrelated to case status and another when one variable related to case status is left out. As in the previous simulations, the failure times were generated by (1), censoring was exponential with mean 5, failure indicators were set to be missing completely at random or missing at random,  $X_{i2} \sim N(\Delta_i, 0.3)$  and  $Q_i = I(X_{i2} > 0.5)$  for  $i = 1, \dots, n$ . We also generated  $X_{i3} \sim N(0, 1)$  where  $X_{i1}, X_{i2}, X_{i3}$  were mutually independent.

In the previous simulations, we fit the data to (11) with  $X_i = \{X_{i1}, X_{i2}\}$  where  $X_{i1} \sim N(2, 1)$ . The additional simulations instead used the covariates and parameters as follows:

- (1)  $\tilde{X}_i = \{1, X_{i1}, X_{i2}, X_{i3}\}$ ,  $\tilde{\alpha} = \{\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3\}$
- (2)  $\tilde{X}_i = \{1, X_{i1}\}$ ,  $\tilde{\alpha} = \{\tilde{\alpha}_0, \tilde{\alpha}_1\}$

That is, rather than fitting model (11) to the data, we modeled the case probability with

$$\text{logit}\{Pr(\Delta_i = 1)\} = \tilde{\alpha}'\tilde{X}_i. \quad (12)$$

Results, shown in Appendix A, remained similar under both alternative models. This indicates that the proposed methods are robust to alternative specification of the logistic regression model.

#### 4. Data Application

In this section, we apply our method to time-to-event data in the Orofacial Pain Prospective Evaluation and Risk Assessment (OPPERA) study. OPFERA is a prospective cohort study designed to identify risk factors for first-onset TMD. A total of 3,263 initially TMD-free individuals were enrolled in OPFERA, which were recruited at four study sites from 2006-

2008. TMD status was confirmed by an RDC examination (Dworkin and LeResche, 1992). For more details on the OPPERA study, see Maixner et al. (2011) and Slade et al. (2011).

Upon enrollment in the study, each OPPERA participant was assessed for psychological distress, previous history of painful conditions, and sensitivity to experimental pain. See Ohrbach et al. (2011), Fillingim et al. (2011), Greenspan et al. (2011), and Maixner et al. (2011) for a complete description of the baseline data sets that were collected in OPPERA. After enrollment each participant was instructed to complete a survey (called a quarterly health update or QHU) every three months. This questionnaire evaluates the frequency and severity of pain in the orofacial region during the previous three months. The QHU also evaluates pain experienced in other bodily regions, use of medications, sleep quality, psychological distress, and other possible risk factors for TMD. For a complete description of the QHU, see Slade et al. (2012).

Participants were screened as positive or negative based on whether or not their QHU responses were indicative of TMD. Participants experiencing at least 2 months with 5 or more days of orofacial pain and at least one day of pain within the past two weeks or with 5 or more days of pain in the last two weeks were considered positive on the QHU. Participants who screened positive on a QHU were asked to undergo a follow-up RDC examination by an expert dentist to determine whether or not they would be classified as an incident case of TMD.

Of the 3,263 subjects, 2,737 filled out at least 1 QHU, and the remaining 521 did not fill out any QHUs. The total number of QHUs was 28,400, of which 26,666 corresponded to subjects not yet diagnosed with TMD. (Subjects diagnosed with TMD continued to complete QHU's, but since these additional QHUs were not relevant to our time-to-event analysis, we did not analyze them.)

There were 1,157 positive QHUs, of which 717 were from individuals free of TMD at the

time of their QHUs. Among these positive QHU's, 486 (about 68%) received an RDC exam. In addition, upon finding a new incident case, investigators asked a matched control to come in for a clinical examination. Controls were matched on study site, gender, and enrollment date within 2 weeks. In this manner, RDC exams were conducted on 342 matched controls. This protocol resulted in 260 positive diagnoses of TMD and 556 negative diagnoses of TMD.

One of the examiners was deemed to be unreliable, so we set all of her RDC exam results to be missing and imputed them using the methods in this paper. This left 404 positive QHUs (56%) resulting in valid clinical exams. We then had 177 positive diagnoses of TMD and 530 negative diagnoses. On the individual level, after setting the exams from the unreliable examiner to missing, there were 514 people who had at least one positive QHU, 401 of which triggered positive only once, and only 39 of which had more than 2 positive QHUs.

#### 4.1 Hazard Ratios

We applied our method to the OPPERA cohort to adjust for participants with missing RDC examinations. First, we predict via logistic regression the probability of a subject being diagnosed as an incident case of TMD given a positive screening on a QHU. Due to the rich body of information collected in each QHU, we carefully selected a small number of predictor variables. Specifically, we fit a generalized linear mixed model with a logistic link function to predict the result of the RDC exam based on each item in the QHU. A mixed model was used because a sizeable ( $n=113$ ) minority of subjects screened positive on multiple QHUs. All models were adjusted for study site and included a random effect term for each subject.

The majority of the variables measured on the QHU were not associated with the result of the RDC exam. The strongest predictor of a positive RDC exam was a count of facial symptoms (e.g. stiffness, fatigue, or soreness) in the previous three months. The time elapsed from baseline to the QHU was also predictive of a positive diagnosis of TMD. Several other possible predictors of a positive RDC exam were identified, but including these additional



predictors in the model did not significantly improve the predictive accuracy of the model. Thus, we predicted the probability of a positive RDC exam based on the count of facial symptoms, time since baseline, and study site. This model was used to perform multiple imputation for those who had positive QHUs but did not come in for their RDC exams. These imputed data sets were used to fit a series of Cox proportional hazards models to estimate the hazard ratio (and associated confidence interval and p-value) for each predictor using the methods described in section 2.1. Examples of predictors include perceived stress, history of comorbid chronic pain conditions and smoking status, among others.

In addition, we examined univariate relationships between the act of coming to the clinic for an RDC exam and numerous other factors. Differences between those who came in and those who did not come in were not significant. There were no significant associations between the baseline variables and the acts of filling out the QHUs or returning for the RDC exam. This indicates the missingness is likely random.

Table 2 shows the results of applying our method to a subset of the putative risk factors of TMD measured in OPPERA. Due to the large number of putative risk factors measured in OPPERA, we only report a subset of these in this table. All continuous variables in the table were normalized to have mean 0 and standard deviation 1 prior to fitting the Cox models. (Thus, the hazard ratios for the continuous variables represent the hazard ratios corresponding to a one-standard deviation increase in the predictor variable.) In Table 2, all the quantitative sensory testing (QST) and psychosocial variables were continuous, and all of the clinical variables were dichotomous (and hence were not normalized).

[Table 2 about here.]

Compared to the unimputed results which treated missing censoring indicators as censored observations, imputation appeared to slightly reduce the hazard ratios for most of the psychosocial variables that were measured in OPPERA. For instance, Table 2 shows the

(standardized) hazard ratios for the Pennebaker Inventory of Limbic Languidness (PILL) score, the neuroticism subscale of the Eysenk Personality Questionnaire (EPQ), the Spielberger Trait Anxiety Inventory score, the Perceived Stress Scale, and the somatization subscale of the symptom checklist-90 (SCL90R). (See Fillingim et al. (2011) for a complete description of these psychosocial instruments.) In each case, the hazard ratios were reduced after imputation.

A similar pattern was observed after applying our imputation method to the OPPERA QST data. The mechanical pain aftersensation ratings were strongly associated with first-onset TMD before imputation, but they were only weakly associated with first-onset TMD after imputation. The pressure pain algometer ratings were also more weakly associated with TMD after imputation (and two of three ratings in Table 2 were no longer significantly associated with first-onset TMD at the  $p < 0.05$  level). See Greenspan et al. (2011) for a more detailed description of these QST measures.

Interestingly, the hazard ratios for the presence of one or more palpation tender points at the temporalis and masseter were also attenuated after imputation. See Ohrbach et al. (2011) for a more detailed description of these variables. These tender points were evaluated as part of the RDC examination using a different protocol than the QST algometer pain ratings. However, both pain measures (algometer and palpation) were measured at the same anatomical locations on the jaw. While the palpation ratings were more strongly associated with first-onset TMD than the algometer ratings both before and after imputation, it is interesting that different pain sensitivity measures using different protocols at the same anatomical location were both attenuated by imputation.

The effects of other clinical variables were also attenuated after imputation. For example, the hazard ratios associated with being unable to open one's mouth wide in the past month and having two or more comorbid pain conditions were both noticeably attenuated after

imputation. However, other clinical variables were more strongly associated with first-onset TMD after imputation. For example, having a history of at least one of five respiratory conditions was only weakly associated with first-onset TMD before imputation (HR=1.38,  $p=0.04$ ), but the association was much stronger after imputation (HR=1.43,  $p=0.004$ ). Also, being a current smoker was not significantly associated with first-onset TMD before imputation (HR=1.26,  $p=0.24$ ) but was associated after imputation (HR=1.49,  $p=0.02$ ). See Ohrbach et al. (2011) for a more detailed description of these variables.

#### 4.2 Incidence Rates

Another important aim of the OPPERA study is to calculate incidence rates of first-onset TMD. In table 3, we calculated incidence rates in two ways. First we treated all missing censoring indicators as censored. Second, we implemented the multiple imputation method in this paper. Overall rates with multiple imputation increased by nearly 2/3 compared to the unimputed rates and by 64% for females and 72% for males. Rates for whites and Hispanics were 99% and 193% higher with imputation. Thus, the incidence rate of first-onset TMD is underestimated without imputation.

[Table 3 about here.]

## 5. Discussion

Motivated by the research questions and framework of the OPPERA study, we presented a computationally efficient method to adjust for missing censoring indicators in time-to-event data using logistic regression and multiple imputation. Logistic regression is used to estimate the failure probability for subjects with missing censoring indicators. Then, the values are imputed and standard errors are estimated via multiple imputation. This framework is important in studies where failure status may be unknown, e.g. with interim event adjudication.

Magder and Hughes (1997) used a similar approach to estimate parameters in a logistic regression model given observed covariates when the outcome variable is measured with uncertainty by incorporating information about sensitivity and specificity. Their approach utilizes the EM algorithm to update the predicted probabilities and parameter estimates and iterating until convergence. Our assumption of MAR data renders iteration unnecessary. In fact, we conducted additional simulations, which found that our one-step procedure of estimating the predicted probabilities was nearly equivalent to the iterative procedure.

In all of the simulations, multiple imputation was the most desirable method, with the narrowest valid confidence intervals and no significant bias. In particular, the method of Cook and Kosorok (2004) produced slightly wider confidence intervals in each simulation we considered. It is unclear if this finding is true in general, perhaps due to a slight bias-variance trade-off or if it is specific to the simulation scenarios we considered. In any event, the differences were extremely small, so the performance of the two methods appears to be comparable for most practical purposes.

However, we believe our method nevertheless has several possible advantages over the method of Cook and Kosorok (2004). First, we programmed their variance estimator and did not get reasonable results for the OPPERA study. Secondly, bootstrapping is much more intensive computationally than our multiple imputation approach. Calculating bootstrap confidence intervals generally requires at least 1000 bootstrap replicates (Efron and Tibshirani, 1993), whereas as few as 10 imputed data sets may be sufficient for multiple imputation (Little and Rubin, 2002). Although the difference in the computing time of the two methods is small for a single fitted model, many such models will need to be fitted in the course of the OPPERA study. OPPERA has already collected data on approximately 3000 genetic markers and has plans to collect data on approximately a million genetic markers in a genome-wide association study. Thus, at least 3000 (and potentially as many as a million) Cox

models will need to be fit, and our proposed method may allow for a significant decrease in computing time. Moreover, our method can also be easily implemented in popular statistical software packages (such as SAS) without additional programming.

The assumption of missing at random seems reasonable for the OPPERA study. However, our methodology can be generalized for non-ignorable missing censoring indicators by using the EM-algorithm to fit a logistic regression model, estimate predicted probabilities, refit the data to a model weighted by the predicted probabilities and iterating until convergence. This is a topic of future study.

Additionally, our methodology may easily be extended to other survival models, such as Poisson regression. We conducted simulations (Table 12 in the Appendix) that produced similar results for Poisson regression compared to the Cox model. In particular, estimates for failure rates were biased when missing censoring indicators were treated as censored and unbiased when we employed the methodology in this paper.

In the OPPERA dataset, hazard ratios, confidence intervals, and p-values changed noticeably for a number of variables. Although some of the results remained qualitatively unchanged based on whether or not our methodology was utilized, we note that even small changes in hazard ratios are important. In addition, estimated incidence rates were significantly increased after imputation. Since the results of OPPERA may become normative in the orofacial pain literature, precise calculation of the incidence rate of TMD and the hazard ratios associated with putative risk factors is important. Thus, imputation is recommended.

#### ACKNOWLEDGEMENTS

The authors would like to acknowledge and thank the principal investigators of the OPPERA study, namely William Maixner, Luda Diatchenko, Bruce Weir, Richard Ohrbach, Roger Fillingim, Joel Greenspan, and Ronald Dubner. The OPPERA study was supported by NIH/NIDCR grant U01DE017018.

## REFERENCES

- Bender, R., Augustin, T., and Blettner, M. (2005). Generating survival times to simulate cox proportional hazards models. *Statistics in Medicine* **24**, 1713–1723.
- Chen, P., He, R., Shen, J.-s., and Sun, J.-g. (2009). Regression analysis of right-censored failure time data with missing censoring indicators. *Acta Mathematicae Applicatae Sinica (English Series)* **25**, 415–426. 10.1007/s10255-008-8807-1.
- Cook, T. D. and Kosorok, M. R. (2004). Analysis of time-to-event data with incomplete event adjudication. *Journal of the American Statistical Association* **99**, pp. 1140–1152.
- Dewanji, A. (1992). A note on a test for competing risks with missing failure type. *Biometrika* **79**, pp. 855–857.
- Dinse, G. E. (1982). Nonparametric estimation for partially-complete time and type of failure data. *Biometrics* **38**, pp. 417–431.
- Dworkin, S. and LeResche, L. (1992). Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. *Journal of Craniomandibular Disorders* **6**, 301–355.
- Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall/CRC, Boca Raton, FL.
- Fillingim, R. B., Ohrbach, R., Greenspan, J. D., Knott, C., Dubner, R., Bair, E., Baraian, C., Slade, G. D., and Maixner, W. (2011). Potential psychosocial risk factors for chronic tmd: Descriptive data and empirically identified domains from the opera case-control study. *The Journal of Pain* **12**, T46 – T60.
- Flehinger, B. J., Reiser, B., and Yashchin, E. (1998). Survival with competing risks and masked causes of failures. *Biometrika* **85**, pp. 151–164.
- Gao, G. and Tsiatis, A. A. (2005). Semiparametric estimators for the regression coefficients in the linear transformation competing risks model with missing cause of failure. *Biometrika*

- 92**, pp. 875–891.
- Gijbels, I., Lin, D., and Ying, Z. (2007). Non- and semi- parametric analysis of failure-time data with missing failure indicators. *Lecture Notes-Monograph Series* **54**, 203–223.
- Goetghebeur, E. and Ryan, L. (1990). A modified log rank test for competing risks with missing failure type. *Biometrika* **77**, pp. 207–211.
- Goetghebeur, E. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing. *Biometrika* **82**, 821–833.
- Greenspan, J. D., Slade, G. D., Bair, E., Dubner, R., Fillingim, R. B., Ohrbach, R., Knott, C., Mulkey, F., Rothwell, R., and Maixner, W. (2011). Pain sensitivity risk factors for chronic tmd: Descriptive data and empirically identified domains from the oppera case control study. *The Journal of Pain* **12**, T61 – T74.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley Series in Probability and Mathematical Statistics.
- Magder, L. S. and Hughes, J. P. (1997). Logistic regression when the outcome is measured with uncertainty. *American Journal of Epidemiology* **146**, 195–203.
- Maixner, W., Diatchenko, L., Dubner, R., Fillingim, R. B., Greenspan, J. D., Knott, C., Ohrbach, R., Weir, B., and Slade, G. D. (2011). Orofacial pain prospective evaluation and risk assessment study the oppera study. *The Journal of Pain* **12**, T4–T11.
- Maixner, W., Greenspan, J. D., Dubner, R., Bair, E., Mulkey, F., Miller, V., Knott, C., Slade, G. D., Ohrbach, R., Diatchenko, L., and Fillingim, R. B. (2011). Potential autonomic risk factors for chronic tmd: Descriptive data and empirically identified domains from the oppera case-control study. *The Journal of Pain* **12**, T75 – T91.
- McKeague, I. W. and Subramanian, S. (1998). Product-limit estimators and cox regression with missing censoring information. *Scandinavian Journal of Statistics* **25**, pp. 589–601.
- Ohrbach, R., Fillingim, R. B., Mulkey, F., Gonzalez, Y., Gordon, S., Gremillion, H., Lim,

- P.-F., Ribeiro-Dasilva, M., Greenspan, J. D., Knott, C., Maixner, W., and Slade, G. (2011). Clinical findings and pain symptoms as potential risk factors for chronic tmd: Descriptive data and empirically identified domains from the oppera case-control study. *The Journal of Pain* **12**, T27 – T45.
- Racine-Poon, A. H. and Hoel, D. G. (1984). Nonparametric estimation of the survival function when cause of death is uncertain. *Biometrics* **40**, pp. 1151–1158.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, pp. 581–592.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association* **91**, pp. 473–489.
- Slade, G. D., Bair, E., By, K., Mulkey, F., Baraian, C., Rothwell, R., Reynolds, M., Miller, V., Gonzalez, Y., Gordon, S., Ribeiro-Dasilva, M., Lim, P. F., Greenspan, J. D., Dubner, R., Fillingim, R. B., Diatchenko, L., Maixner, W., Dampier, D., Knott, C., and Ohrbach, R. (2011). Study methods, recruitment, sociodemographic findings, and demographic representativeness in the oppera study. *The Journal of Pain* **12**, T12–T26.
- Slade, G. D., Sanders, A., Bair, E., Brownstein, N. C., Fillingim, R. B., Maixner, W., Greenspan, J. D., and Ohrbach, R. (2012). Pre-clinical episodes of orofacial pain symptoms and their association with healthcare behaviors in the oppera prospective cohort study. Submitted.
- Snapinn, S. M. (1998). Survival analysis with uncertain endpoints. *Biometrics* **54**, pp. 209–218.
- Subramanian, S. (2000). Efficient estimation of regression coefficients and baseline hazard under proportionality of conditional hazards. *Journal of Statistical Planning and Inference* **84**, 81 – 94.
- Subramanian, S. (2004). Asymptotically efficient estimation of a survival function in the missing censoring indicator model. *Journal of Nonparametric Statistics* **16**, 797–817.



- Subramanian, S. (2006). Survival analysis for the missing censoring indicator model using kernel density estimation techniques. *Statistical Methodology* **3**, 125 – 136.
- Subramanian, S. (2009). The multiple imputations based kaplan-meier estimator. *Statistics & Probability Letters* **79**, 1906 – 1914.
- Subramanian, S. (2011). Multiple imputations and the missing censoring indicator model. *Journal of Multivariate Analysis* **102**, 105 – 117.
- Subramanian, S. and Bandyopadhyay, D. (2010). Doubly robust semiparametric estimation for the missing censoring indicator model. *Statistics & Probability Letters* **80**, 621 – 630.
- Subramanian, S. and Bean, D. (2008a). Hazard function estimation from homogeneous right censored data with missing censoring indicators. *Statistical Methodology* **5**, 515 – 527.
- Subramanian, S. and Bean, D. (2008b). The missing censoring indicator model and the smoothed bootstrap. *Computational Statistics & Data Analysis* **53**, 471 – 476.
- Tsiatis, A. A., Davidian, M., and McMeney, B. (2002). Multiple imputation methods for testing treatment differences in survival distributions with missing cause of failure. *Biometrika* **89**, 238–244.
- van der Laan, M. J. and McKeague, I. W. (1998). Efficient estimation from right-censored data when failure indicators are missing at random. *The Annals of Statistics* **26**, 164–182.



## APPENDIX OVERVIEW OF SIMULATIONS

In this appendix, we provide the results of additional simulations. We investigate the performance of the method under a variety of missing data mechanisms as well as when we specify alternative logistic regression models for the probability of being a case given that the participant screened positive on the simple examination.

Recall that we created missing censoring indicators under the following classical missing data mechanisms of Rubin (1976):

- (I) The probability of having a missing censoring indicator is independent of the data. This is known as missing completely at random (MCAR).
- (II) The probability of having a missing censoring indicator depends on an observed covariate. This is known as missing at random (MAR).
- (III) The probability of having a missing censoring indicator depends on the censoring indicator. This is known as missing not at random (MNAR).

For mechanism I, we randomly set 40% of the censoring indicators,  $\Delta_i = I(T_i < C_i)$ , to be missing. For mechanism III, we investigated two scenarios.

- (A) In the first, we set 30% of the censored observations and 50% of the failures to have missing indicators.
- (B) In the second, we set 20% of the censored observations and 60% of the failures to have missing indicators.

## APPENDIX SIMULATIONS UNDER MCAR

[Table 4 about here.]

When the data were MCAR, our method had less bias on average than the complete case method depended on the true parameter value. Not only did our method have adequate coverage, but it had the most narrow confidence intervals of the methods with adequate

coverage. As in other simulations, the method treating all missing indicators as failures had poor coverage and introduced extreme bias. The complete case method and the method that treat all missing censoring indicators as censored were valid, but had much wider confidence intervals than our method.

However, note that the complete case method would not be applicable to the OPPERA study. According to the OPPERA protocol, participants who do not screen positive on the QHU are automatically considered censored. Only participants who screen positive on the QHU (i.e. those with  $Q = 1$ ) may potentially have missing censoring indicators.

#### APPENDIX ADDITIONAL SIMULATIONS UNDER MAR

In order to more closely parallel the OPPERA study, we simulated data for which we randomly set to missing 40% of the censoring indicators for those with  $Q = 1$ . This setup corresponds to the still strict assumption that the probability that a participant completes an RDC exam depends only on whether or not their QHU was positive. Results are included in Table 5. All methods had a negligible amount of bias in these scenarios except for the complete case method and the method that treated all missing indicators as failures. In these simulations, the complete case method also displayed extreme bias and poor coverage. This indicates that a complete case analysis would not be appropriate for a study such as OPPERA.

[Table 5 about here.]

#### APPENDIX ALTERNATIVE LOGISTIC REGRESSION MODELS

We investigated alternate logistic regression models for the probability of being a case given covariates. Recall that we originally modeled the probability of being a case as

$$p(\Delta_i = 1|X_i, \alpha) = \frac{\exp(\alpha'X_i)}{1 + \exp(\alpha'X_i)} \quad (\text{A.1})$$

The alternative models were of the form (A.1) but used the covariates

(1)  $X_i = \{X_{i1}, X_{i2}, X_{i3}\}$

(2)  $X_i = \{X_{i1}\}$

The original logistic model had the covariate  $X_i = \{X_{i1}, X_{i2}\}$  where  $X_{i1} \sim N(0, 2)$ ,  $X_{i1} \sim N(0, \Delta_i)$ ,  $X_{i3} \sim N(0, 1)$  are mutually independent for  $j = 1, 2, 3$  and  $i = 1, \dots, n$ .

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

[Table 9 about here.]



## APPENDIX SIMULATIONS UNDER MNAR

[Table 10 about here.]

[Table 11 about here.]

When the data were MNAR, bias increased for all methods. In particular, the complete case method consistently displayed a high amount of bias and did not achieve the desired coverage rate. For our imputation method and the method of Cook and Kosorok, bias increased and coverage decreased as the true parameter value increased. This indicates that when the MAR assumption is violated, our method as well as the method of Cook and Kosorok may not be valid. On the other hand, even when the data was not missing at random, our provided an improvement in terms of bias and coverage over the complete case method and the method that treats all missing subjects as failures.

## APPENDIX SIMULATIONS FOR POISSON REGRESSION

We investigated the performance of our method if the desired time-to-event analysis was a Poisson regression model rather than a Cox model. Poisson models are commonly used to estimate incidence rates, which were desired in the OPPERA study. The following section details the simulations.

The simulations were identical to those in the main paper, with the exception of fitting the imputed data to Poisson regression models rather than to Cox proportional hazards models. That is, we fit the data from imputation  $j = 1, \dots, m$  to the model

$$\log(dN_{ij}^{(I)}) = \alpha + \beta x_{i1} + \log(Y_i)$$

We measured the bias, defined as  $\hat{\beta}$  minus the true value, where  $\beta \in -0.5, -1.5, -3$ .

The Cook and Kosorok method does not immediately generalize to Poisson regression. Consequently, we only compared our method to the unachievable ideal of no missing data, the complete case method, and the two ad-hoc methods.

As in previous simulations, our method had coverage close to the nominal rate. None of the other methods had proper coverage for all of the simulations. Multiple imputation yielded the least bias of all the methods besides the unacheivable ideal of observing all data and produced more narrow confidence intervals than the complete case method and the method that treats all missing censoring indicators as censored. In conclusion, our method is the only valid method to estimate incidence rates.

[Table 12 about here.]



**Table 1**  
*Simulation Results for MAR*

| $\beta$ | Method                | Bias    | SE (Bias) | Width  | SE (Width) | Coverage | MC Error |
|---------|-----------------------|---------|-----------|--------|------------|----------|----------|
| -0.5    | Full Data             | 0.0001  | 0.0005    | 0.1668 | 0.0001     | 0.943    | 0.007    |
|         | Complete Case         | -0.0049 | 0.0007    | 0.2213 | 0.0001     | 0.951    | 0.007    |
|         | Treat all as Censored | 0.0963  | 0.0007    | 0.2189 | 0.0001     | 0.566    | 0.007    |
|         | Treat all as Failures | 0.0039  | 0.0006    | 0.1701 | 0.0001     | 0.939    | 0.007    |
|         | Cook & Kosorok        | 0.0004  | 0.0006    | 0.1736 | 0.0001     | 0.945    | 0.007    |
|         | Multiple Imputation   | 0.0004  | 0.0006    | 0.1717 | 0.0001     | 0.937    | 0.007    |
| -1.5    | Full Data             | -0.0048 | 0.0011    | 0.3182 | 0.0002     | 0.930    | 0.007    |
|         | Complete Case         | -0.0786 | 0.0015    | 0.4388 | 0.0004     | 0.897    | 0.007    |
|         | Treat all as Censored | 0.1004  | 0.0014    | 0.4269 | 0.0003     | 0.832    | 0.007    |
|         | Treat all as Failures | 0.0641  | 0.0011    | 0.3158 | 0.0002     | 0.853    | 0.007    |
|         | Cook & Kosorok        | -0.0051 | 0.0011    | 0.3384 | 0.0004     | 0.940    | 0.007    |
|         | Multiple Imputation   | -0.0052 | 0.0011    | 0.3310 | 0.0002     | 0.943    | 0.007    |
| -3      | Full Data             | -0.0301 | 0.0025    | 0.7627 | 0.0009     | 0.957    | 0.007    |
|         | Complete Case         | -0.2212 | 0.0037    | 1.0990 | 0.0017     | 0.903    | 0.007    |
|         | Treat all as Censored | 0.0759  | 0.0035    | 1.0549 | 0.0016     | 0.928    | 0.007    |
|         | Treat all as Failures | 0.5633  | 0.0024    | 0.6365 | 0.0006     | 0.120    | 0.007    |
|         | Cook & Kosorok        | -0.0325 | 0.0027    | 0.8846 | 0.0016     | 0.942    | 0.007    |
|         | Multiple Imputation   | -0.0333 | 0.0027    | 0.8060 | 0.0011     | 0.947    | 0.007    |

\*\* : denotes that the empirical bias (coverage probability) is significantly different from 0 (0.95) with  $p < 0.01$ .

\*\*\* : denotes that the empirical bias (coverage probability) is significantly different from 0 (0.95) with  $p < 0.001$ .



**Table 2**  
Results from the OPFERA Study

| Clinical Variable                                     | Consider All MCIs as Censored |      |      |         | Multiple Imputation |      |      |         |
|---|-------------------------------|------|------|---------|---------------------|------|------|---------|
|   | HR                            | LCL  | UCL  | P       | HR                  | LCL  | UCL  | P       |
| <b>Clinical Variable</b>                              |                               |      |      |         |                     |      |      |         |
| In the last month could not open mouth wide           | 3.26                          | 1.83 | 5.84 | <0.0001 | 2.45                | 1.42 | 4.22 | 0.0012  |
| Has two or more comorbid chronic pain disorders       | 3.08                          | 2.26 | 4.21 | <0.0001 | 2.50                | 1.90 | 3.29 | <0.0001 |
| History of 5 respiratory conditions                   | 1.38                          | 1.01 | 1.87 | 0.0408  | 1.45                | 1.13 | 1.87 | 0.0040  |
| Smoking: current                                      | 1.26                          | 0.86 | 1.84 | 0.2403  | 1.49                | 1.07 | 2.09 | 0.0199  |
| Smoking: former                                       | 1.87                          | 1.22 | 2.87 | 0.0041  | 1.65                | 1.12 | 2.43 | 0.0106  |
| One or more palpation tender points: right temporalis | 1.83                          | 1.32 | 2.52 | 0.0002  | 1.54                | 1.18 | 2.02 | 0.0017  |
| One or more palpation tender points: left temporalis  | 1.60                          | 1.14 | 2.25 | 0.0064  | 1.48                | 1.12 | 1.97 | 0.0060  |
| One or more palpation tender points: right masseter   | 1.85                          | 1.35 | 2.53 | 0.0001  | 1.63                | 1.25 | 2.12 | 0.0003  |
| One or more palpation tender points: left masseter    | 1.70                          | 1.23 | 2.35 | 0.0013  | 1.53                | 1.17 | 2.01 | 0.0021  |
| <b>Quantitative Sensory Testing Variable</b>          |                               |      |      |         |                     |      |      |         |
| Pressure pain threshold: temporalis                   | 1.26                          | 1.07 | 1.49 | 0.0065  | 1.16                | 1.01 | 1.33 | 0.0335  |
| Pressure pain threshold: masseter                     | 1.23                          | 1.04 | 1.45 | 0.0170  | 1.15                | 1.00 | 1.32 | 0.0576  |
| Pressure pain threshold: TM joint                     | 1.25                          | 1.05 | 1.48 | 0.0106  | 1.14                | 1.00 | 1.30 | 0.0555  |
| Mechanical pain aftersensation: 512mN probe, 15 s     | 1.23                          | 1.09 | 1.38 | 0.0006  | 1.15                | 1.03 | 1.28 | 0.0123  |
| Mechanical pain aftersensation: 512mN probe, 30 s     | 1.20                          | 1.07 | 1.34 | 0.0020  | 1.12                | 1.01 | 1.25 | 0.0328  |
| <b>Psychosocial Variable</b>                          |                               |      |      |         |                     |      |      |         |
| PILL Global Score                                     | 1.52                          | 1.35 | 1.71 | <0.0001 | 1.46                | 1.31 | 1.62 | <0.0001 |
| EPQ-R Neuroticism                                     | 1.39                          | 1.21 | 1.60 | <0.0001 | 1.26                | 1.12 | 1.42 | 0.0002  |
| Trait Anxiety Inventory                               | 1.43                          | 1.25 | 1.64 | <0.0001 | 1.35                | 1.21 | 1.52 | <0.0001 |
| Perceived Stress Scale                                | 1.35                          | 1.17 | 1.55 | <0.0001 | 1.30                | 1.16 | 1.47 | <0.0001 |
| SCL 90R Somatization                                  | 1.44                          | 1.31 | 1.58 | <0.0001 | 1.40                | 1.29 | 1.52 | <0.0001 |





**Table 3**  
*Estimated TMD Incidence Rates (in Percentages) With and Without Imputation*

|          | No MI | MI   | Percent Change |
|----------|-------|------|----------------|
| Overall  | 2.23  | 3.70 | 66             |
| Males    | 1.87  | 3.22 | 72             |
| Females  | 2.46  | 4.03 | 64             |
| White    | 1.70  | 3.37 | 99             |
| Black    | 4.20  | 5.32 | 27             |
| Hispanic | 1.17  | 3.44 | 193            |
| Other    | 1.10  | 1.80 | 63             |



**Table 4**  
*Simulation Results for MCAR*

| $\beta$ | Method                | Bias    | SE (Bias) | Width  | SE (Width) | Coverage | MC Error |
|---------|-----------------------|---------|-----------|--------|------------|----------|----------|
| -0.5    | Full Data             | -0.0010 | 0.0006    | 0.1671 | 0.0001     | 0.930    | 0.007    |
|         | Complete Case         | -0.0042 | 0.0007    | 0.2218 | 0.0001     | 0.956    | 0.007    |
|         | Treat all as Censored | -0.0021 | 0.0007    | 0.2212 | 0.0001     | 0.947    | 0.007    |
|         | Treat all as Failures | 0.0932  | 0.0005    | 0.1525 | 0.0001     | 0.323    | 0.007    |
|         | Cook & Kosorok        | -0.0005 | 0.0006    | 0.1784 | 0.0002     | 0.939    | 0.007    |
|         | Multiple Imputation   | -0.0006 | 0.0006    | 0.1714 | 0.0001     | 0.935    | 0.007    |
| -1.5    | Full Data             | -0.0008 | 0.0010    | 0.3185 | 0.0002     | 0.966    | 0.007    |
|         | Complete Case         | -0.0056 | 0.0014    | 0.4261 | 0.0003     | 0.948    | 0.007    |
|         | Treat all as Censored | -0.0025 | 0.0014    | 0.4229 | 0.0003     | 0.950    | 0.007    |
|         | Treat all as Failures | 0.8277  | 0.0007    | 0.2036 | 0.0001     | 0.000    | 0.007    |
|         | Cook & Kosorok        | 0.0002  | 0.0011    | 0.3455 | 0.0004     | 0.946    | 0.007    |
|         | Multiple Imputation   | 0.0000  | 0.0011    | 0.3307 | 0.0002     | 0.951    | 0.007    |
| -3      | Full Data             | -0.0190 | 0.0025    | 0.7574 | 0.0008     | 0.952    | 0.007    |
|         | Complete Case         | -0.0321 | 0.0036    | 1.0255 | 0.0016     | 0.942    | 0.007    |
|         | Treat all as Censored | -0.0205 | 0.0035    | 1.0070 | 0.0015     | 0.950    | 0.007    |
|         | Treat all as Failures | 2.5225  | 0.0009    | 0.2216 | 0.0001     | 0.000    | 0.007    |
|         | Cook & Kosorok        | -0.0199 | 0.0027    | 0.9029 | 0.0020     | 0.949    | 0.007    |
|         | Multiple Imputation   | -0.0210 | 0.0028    | 0.7951 | 0.0010     | 0.936    | 0.007    |

**Table 5**  
*Additional Simulation Results for MAR*

| $\beta$ | Method                | Bias    | SE (Bias) | Width  | SE (Width) | Coverage | MC Error |
|---------|-----------------------|---------|-----------|--------|------------|----------|----------|
| -0.5    | Full Data             | -0.0010 | 0.0006    | 0.1671 | 0.0001     | 0.930    | 0.007    |
|         | Complete Case         | -0.0543 | 0.0008    | 0.2210 | 0.0001     | 0.825    | 0.007    |
|         | Treat all as Censored | -0.0021 | 0.0007    | 0.2212 | 0.0001     | 0.947    | 0.007    |
|         | Treat all as Failures | 0.0041  | 0.0006    | 0.1703 | 0.0001     | 0.929    | 0.007    |
|         | Cook & Kosorok        | -0.0012 | 0.0006    | 0.1742 | 0.0001     | 0.934    | 0.007    |
|         | Multiple Imputation   | -0.0012 | 0.0006    | 0.1722 | 0.0001     | 0.938    | 0.007    |
| -1.5    | Full Data             | -0.0008 | 0.0010    | 0.3185 | 0.0002     | 0.966    | 0.007    |
|         | Complete Case         | -0.1329 | 0.0014    | 0.4283 | 0.0003     | 0.759    | 0.007    |
|         | Treat all as Censored | -0.0025 | 0.0014    | 0.4229 | 0.0003     | 0.950    | 0.007    |
|         | Treat all as Failures | 0.0849  | 0.0011    | 0.3147 | 0.0002     | 0.790    | 0.007    |
|         | Cook & Kosorok        | -0.0006 | 0.0011    | 0.3411 | 0.0004     | 0.951    | 0.007    |
|         | Multiple Imputation   | -0.0006 | 0.0011    | 0.332  | 0.0002     | 0.957    | 0.007    |
| -3      | Full Data             | -0.0190 | 0.0025    | 0.7574 | 0.0008     | 0.952    | 0.007    |
|         | Complete Case         | -0.2342 | 0.0037    | 1.0336 | 0.0016     | 0.883    | 0.007    |
|         | Treat all as Censored | -0.0205 | 0.0035    | 1.0070 | 0.0015     | 0.950    | 0.007    |
|         | Treat all as Failures | 0.6626  | 0.0025    | 0.6165 | 0.0006     | 0.047    | 0.007    |
|         | Cook & Kosorok        | -0.0232 | 0.0028    | 0.8775 | 0.0016     | 0.940    | 0.007    |
|         | Multiple Imputation   | -0.0240 | 0.0028    | 0.7996 | 0.0010     | 0.937    | 0.007    |

**Table 6**  
*Alternative model of type (1), MCAR*

| $\beta$ | Method                | Bias    | SE (Bias) | Width  | SE (Width) | Coverage |
|---------|-----------------------|---------|-----------|--------|------------|----------|
| -0.5    | Full Data             | -0.0011 | 0.0013    | 0.1670 | 0.0002     | 0.949    |
|         | Complete Case         | -0.0018 | 0.0017    | 0.2162 | 0.0003     | 0.953    |
|         | Treat all as Censored | -0.0009 | 0.0017    | 0.2156 | 0.0002     | 0.962    |
|         | Treat all as Failures | 0.0904  | 0.0012    | 0.1507 | 0.0001     | 0.338    |
|         | Cook & Kosorok        | -0.0012 | 0.0014    | 0.1818 | 0.0003     | 0.953    |
|         | Multiple Imputation   | -0.0013 | 0.0014    | 0.1764 | 0.0002     | 0.947    |
| -1.5    | Full Data             | -0.0039 | 0.0026    | 0.3187 | 0.0005     | 0.957    |
|         | Complete Case         | -0.0031 | 0.0034    | 0.4143 | 0.0009     | 0.952    |
|         | Treat all as Censored | -0.0028 | 0.0034    | 0.4121 | 0.0008     | 0.947    |
|         | Treat all as Failures | 0.8141  | 0.0017    | 0.2026 | 0.0002     | 0        |
|         | Cook & Kosorok        | -0.0042 | 0.0029    | 0.3596 | 0.0010     | 0.937    |
|         | Multiple Imputation   | -0.0048 | 0.0029    | 0.3499 | 0.0007     | 0.947    |
| -3      | Full Data             | -0.0146 | 0.0061    | 0.7586 | 0.0021     | 0.943    |
|         | Complete Case         | -0.0240 | 0.0081    | 0.9998 | 0.0036     | 0.956    |
|         | Treat all as Censored | -0.0241 | 0.0081    | 0.9859 | 0.0032     | 0.957    |
|         | Treat all as Failures | 2.5119  | 0.0021    | 0.2206 | 0.0002     | 0        |
|         | Cook & Kosorok        | -0.0224 | 0.0067    | 0.8779 | 0.0037     | 0.934    |
|         | Multiple Imputation   | -0.0264 | 0.0068    | 0.8442 | 0.0027     | 0.952    |



**Table 7**  
*Alternative model of type (1), MAR*

| $\beta$ | Method                | Bias    | SE (Bias) | Width  | SE (Width) | Coverage |       |
|---------|-----------------------|---------|-----------|--------|------------|----------|-------|
| -0.5    | Full Data             | 0.0002  | 0.0006    | 0.1669 | 0.0001     | 0.939    | 0.007 |
|         | Complete Case         | -0.0033 | 0.0008    | 0.2212 | 0.0001     | 0.936    | 0.007 |
|         | Treat all as Censored | 0.0987  | 0.0007    | 0.2188 | 0.0001     | 0.568    | 0.007 |
|         | Treat all as Failures | 0.0038  | 0.0006    | 0.1702 | 0.0001     | 0.935    | 0.007 |
|         | Cook & Kosorok        | 0.0006  | 0.0006    | 0.1739 | 0.0001     | 0.940    | 0.007 |
|         | Multiple Imputation   | 0.0006  | 0.0006    | 0.1718 | 0.0001     | 0.937    | 0.007 |
| -1.5    | Full Data             | -0.0092 | 0.0010    | 0.3195 | 0.0002     | 0.954    | 0.007 |
|         | Complete Case         | -0.0805 | 0.0015    | 0.4388 | 0.0004     | 0.890    | 0.007 |
|         | Treat all as Censored | 0.0958  | 0.0014    | 0.4275 | 0.0003     | 0.848    | 0.007 |
|         | Treat all as Failures | 0.059   | 0.0010    | 0.3173 | 0.0002     | 0.883    | 0.007 |
|         | Cook & Kosorok        | -0.0092 | 0.0011    | 0.3407 | 0.0004     | 0.944    | 0.007 |
|         | Multiple Imputation   | -0.0094 | 0.0011    | 0.3323 | 0.0002     | 0.947    | 0.007 |
| -3      | Full Data             | -0.0157 | 0.0026    | 0.7559 | 0.0008     | 0.942    | 0.007 |
|         | Complete Case         | -0.2083 | 0.0039    | 1.0920 | 0.0017     | 0.905    | 0.007 |
|         | Treat all as Censored | 0.0929  | 0.0036    | 1.0463 | 0.0015     | 0.924    | 0.007 |
|         | Treat all as Failures | 0.5830  | 0.0025    | 0.6307 | 0.0006     | 0.101    | 0.007 |
|         | Cook & Kosorok        | -0.0165 | 0.0028    | 0.8851 | 0.0016     | 0.940    | 0.007 |
|         | Multiple Imputation   | -0.0161 | 0.0028    | 0.7986 | 0.0010     | 0.941    | 0.007 |



**Table 8**  
*Alternative model of type (2), MCAR*

| $\beta$ | Method                | Bias    | SE (Bias) | Width  | SE (Width) | Coverage | MC Error |
|---------|-----------------------|---------|-----------|--------|------------|----------|----------|
| -0.5    | Full Data             | -0.0022 | 0.0005    | 0.1669 | 0.0001     | 0.959    | 0.007    |
|         | Complete Case         | -0.0554 | 0.0007    | 0.2209 | 0.0001     | 0.845    | 0.007    |
|         | Treat all as Censored | -0.0025 | 0.0007    | 0.2210 | 0.0001     | 0.943    | 0.007    |
|         | Treat all as Failures | 0.0037  | 0.0005    | 0.1701 | 0.0001     | 0.952    | 0.007    |
|         | Cook & Kosorok        | -0.0022 | 0.0006    | 0.1740 | 0.0001     | 0.953    | 0.007    |
|         | Multiple Imputation   | -0.0021 | 0.0006    | 0.1723 | 0.0001     | 0.956    | 0.007    |
| -1.5    | Full Data             | -0.0045 | 0.0010    | 0.3183 | 0.0002     | 0.961    | 0.007    |
|         | Complete Case         | -0.1311 | 0.0013    | 0.4253 | 0.0003     | 0.794    | 0.007    |
|         | Treat all as Censored | -0.0034 | 0.0013    | 0.4209 | 0.0003     | 0.963    | 0.007    |
|         | Treat all as Failures | 0.0810  | 0.0010    | 0.3145 | 0.0002     | 0.818    | 0.007    |
|         | Cook & Kosorok        | -0.0038 | 0.0011    | 0.3478 | 0.0004     | 0.964    | 0.007    |
|         | Multiple Imputation   | -0.0041 | 0.0011    | 0.3379 | 0.0002     | 0.965    | 0.007    |
| -3      | Full Data             | -0.0190 | 0.0025    | 0.758  | 0.0008     | 0.952    | 0.007    |
|         | Complete Case         | -0.2373 | 0.0035    | 1.0341 | 0.0016     | 0.871    | 0.007    |
|         | Treat all as Censored | -0.0250 | 0.0034    | 1.0082 | 0.0014     | 0.947    | 0.007    |
|         | Treat all as Failures | 0.6623  | 0.0025    | 0.6172 | 0.0006     | 0.052    | 0.007    |
|         | Cook & Kosorok        | -0.0240 | 0.0028    | 0.8803 | 0.0016     | 0.935    | 0.007    |
|         | Multiple Imputation   | -0.0264 | 0.0028    | 0.8242 | 0.0010     | 0.945    | 0.007    |



**Table 9**  
*Alternative model of type (2), MAR*

| $\beta$ | Method                | Bias    | SE (Bias) | Width  | SE (Width) | Coverage | MC Error |
|---------|-----------------------|---------|-----------|--------|------------|----------|----------|
| -0.5    | Full Data             | -0.0043 | 0.0006    | 0.1668 | 0.00014    | 0.934    | 0.007    |
|         | Complete Case         | -0.0101 | 0.0007    | 0.2209 | 0.0001     | 0.951    | 0.007    |
|         | Treat all as Censored | 0.0914  | 0.0007    | 0.2186 | 0.0001     | 0.617    | 0.007    |
|         | Treat all as Failures | -0.0013 | 0.0006    | 0.1700 | 0.0001     | 0.949    | 0.007    |
|         | Cook & Kosorok        | -0.0047 | 0.0006    | 0.1741 | 0.0001     | 0.934    | 0.007    |
|         | Multiple Imputation   | -0.0047 | 0.0006    | 0.1719 | 0.0001     | 0.950    | 0.007    |
| -1.5    | Full Data             | -0.0019 | 0.0010    | 0.3186 | 0.0002     | 0.951    | 0.007    |
|         | Complete Case         | -0.0765 | 0.0015    | 0.4391 | 0.0004     | 0.908    | 0.007    |
|         | Treat all as Censored | 0.1009  | 0.0014    | 0.4273 | 0.0003     | 0.841    | 0.007    |
|         | Treat all as Failures | 0.0643  | 0.0011    | 0.3165 | 0.0002     | 0.861    | 0.007    |
|         | Cook & Kosorok        | -0.0043 | 0.0011    | 0.3446 | 0.0004     | 0.947    | 0.007    |
|         | Multiple Imputation   | -0.0046 | 0.0011    | 0.3368 | 0.0002     | 0.944    | 0.007    |
| -3      | Full Data             | -0.0296 | 0.0026    | 0.7606 | 0.0009     | 0.948    | 0.007    |
|         | Complete Case         | -0.2178 | 0.0038    | 1.1013 | 0.0018     | 0.903    | 0.007    |
|         | Treat all as Censored | 0.0827  | 0.0035    | 1.0555 | 0.0015     | 0.933    | 0.007    |
|         | Treat all as Failures | 0.5783  | 0.0025    | 0.6334 | 0.0006     | 0.097    | 0.007    |
|         | Cook & Kosorok        | -0.0294 | 0.0028    | 0.8812 | 0.0016     | 0.937    | 0.007    |
|         | Multiple Imputation   | -0.0314 | 0.0028    | 0.8269 | 0.0011     | 0.951    | 0.007    |



**Table 10**  
*Simulation Results for MNAR, scenario A*

| $\beta$ | Method                | Bias    | SE (Bias) | Width  | SE (Width) | Coverage | MC Error |
|---------|-----------------------|---------|-----------|--------|------------|----------|----------|
| -0.5    | Full Data             | -0.0021 | 0.0006    | 0.1671 | 0.0001     | 0.938    | 0.007    |
|         | Complete Case         | -0.0755 | 0.0008    | 0.2424 | 0.0004     | 0.778    | 0.007    |
|         | Treat all as Censored | -0.0022 | 0.0008    | 0.2421 | 0.0001     | 0.942    | 0.007    |
|         | Treat all as Failures | 0.0023  | 0.0006    | 0.1705 | 0.0001     | 0.942    | 0.007    |
|         | Cook & Kosorok        | -0.0054 | 0.0006    | 0.1760 | 0.0001     | 0.940    | 0.007    |
|         | Multiple Imputation   | -0.0054 | 0.0006    | 0.1732 | 0.0001     | 0.943    | 0.007    |
| -1.5    | Full Data             | -0.0030 | 0.0011    | 0.3185 | 0.0002     | 0.942    | 0.007    |
|         | Complete Case         | -0.1744 | 0.0016    | 0.4691 | 0.0004     | 0.717    | 0.007    |
|         | Treat all as Censored | -0.0049 | 0.0015    | 0.4623 | 0.0004     | 0.947    | 0.007    |
|         | Treat all as Failures | 0.0646  | 0.0011    | 0.3172 | 0.0002     | 0.875    | 0.007    |
|         | Cook & Kosorok        | -0.0206 | 0.0011    | 0.346  | 0.0004     | 0.921    | 0.007    |
|         | Multiple Imputation   | -0.0207 | 0.0011    | 0.3362 | 0.0002     | 0.939    | 0.007    |
| -3      | Full Data             | -0.0289 | 0.0026    | 0.7611 | 0.0009     | 0.948    | 0.007    |
|         | Complete Case         | -0.3308 | 0.0040    | 1.1490 | 0.0018     | 0.824    | 0.007    |
|         | Treat all as Censored | -0.0339 | 0.0039    | 1.1060 | 0.0016     | 0.935    | 0.007    |
|         | Treat all as Failures | 0.5242  | 0.0026    | 0.6487 | 0.0007     | 0.181    | 0.007    |
|         | Cook & Kosorok        | -0.0737 | 0.0029    | 0.9026 | 0.0018     | 0.898    | 0.007    |
|         | Multiple Imputation   | -0.0748 | 0.0029    | 0.8195 | 0.0011     | 0.938    | 0.007    |





**Table 11**  
*Simulation Results for MNAR, scenario B*

|      |                       |         |        |        |        |       |       |
|------|-----------------------|---------|--------|--------|--------|-------|-------|
| -0.5 | Full Data             | -0.0021 | 0.0006 | 0.1671 | 0.0001 | 0.938 | 0.007 |
|      | Complete Case         | -0.0998 | 0.0009 | 0.2713 | 0.0002 | 0.687 | 0.007 |
|      | Treat all as Censored | -0.0016 | 0.0009 | 0.2707 | 0.0001 | 0.943 | 0.007 |
|      | Treat all as Failures | 0.0009  | 0.0006 | 0.1707 | 0.0001 | 0.943 | 0.007 |
|      | Cook & Kosorok        | -0.0094 | 0.0006 | 0.1783 | 0.0001 | 0.933 | 0.007 |
|      | Multiple Imputation   | -0.0094 | 0.0000 | 0.1747 | 0.0001 | 0.934 | 0.007 |
| -1.5 | Full Data             | -0.0030 | 0.0011 | 0.3185 | 0.0002 | 0.942 | 0.007 |
|      | Complete Case         | -0.2278 | 0.0018 | 0.5289 | 0.0005 | 0.618 | 0.007 |
|      | Treat all as Censored | -0.0046 | 0.0017 | 0.5169 | 0.0004 | 0.958 | 0.007 |
|      | Treat all as Failures | 0.0444  | 0.0011 | 0.3201 | 0.0002 | 0.910 | 0.007 |
|      | Cook & Kosorok        | -0.0398 | 0.0012 | 0.3534 | 0.0004 | 0.902 | 0.007 |
|      | Multiple Imputation   | -0.0399 | 0.0012 | 0.3417 | 0.0002 | 0.922 | 0.007 |
| -3   | Full Data             | -0.0289 | 0.0026 | 0.7611 | 0.0009 | 0.948 | 0.007 |
|      | Complete Case         | -0.4316 | 0.0046 | 1.3085 | 0.0023 | 0.781 | 0.007 |
|      | Treat all as Censored | -0.0464 | 0.0044 | 1.2434 | 0.002  | 0.932 | 0.007 |
|      | Treat all as Failures | 0.3661  | 0.0027 | 0.6850 | 0.0007 | 0.451 | 0.007 |
|      | Cook & Kosorok        | -0.1180 | 0.0030 | 0.9425 | 0.0021 | 0.870 | 0.007 |
|      | Multiple Imputation   | -0.1189 | 0.0030 | 0.8395 | 0.0012 | 0.928 | 0.007 |



**Table 12**  
*Simulation Results for Poisson Models, MAR*

| $\beta$ | Method                | Bias    | SE (Bias) | Width  | SE (Width) | Coverage | MC Error |
|---------|-----------------------|---------|-----------|--------|------------|----------|----------|
| -0.5    | Full Data             | -0.0011 | 0.0006    | 0.1593 | 0.0001     | 0.945    | 0.007    |
|         | Complete Case         | -0.0051 | 0.0008    | 0.2112 | 0.0001     | 0.946    | 0.007    |
|         | Treat all as Censored | 0.0963  | 0.0008    | 0.2090 | 0.0001     | 0.569    | 0.007    |
|         | Treat all as Failures | 0.0021  | 0.0006    | 0.1624 | 0.0001     | 0.948    | 0.007    |
|         | Multiple Imputation   | -0.0010 | 0.0006    | 0.1639 | 0.0001     | 0.949    | 0.007    |
| -1.5    | Full Data             | -0.0029 | 0.0010    | 0.2867 | 0.0002     | 0.951    | 0.007    |
|         | Complete Case         | -0.0786 | 0.0014    | 0.3996 | 0.0004     | 0.887    | 0.007    |
|         | Treat all as Censored | 0.1033  | 0.0013    | 0.3876 | 0.0003     | 0.815    | 0.007    |
|         | Treat all as Failures | 0.0617  | 0.0010    | 0.2866 | 0.0002     | 0.845    | 0.007    |
|         | Multiple Imputation   | -0.0040 | 0.0011    | 0.2981 | 0.0002     | 0.959    | 0.007    |
| -3      | Full Data             | -0.0135 | 0.0022    | 0.5984 | 0.0008     | 0.940    | 0.007    |
|         | Complete Case         | -0.2086 | 0.0033    | 0.8727 | 0.0015     | 0.850    | 0.007    |
|         | Treat all as Censored | 0.1062  | 0.0031    | 0.8361 | 0.0012     | 0.904    | 0.007    |
|         | Treat all as Failures | 0.4954  | 0.0025    | 0.5568 | 0.0006     | 0.120    | 0.007    |
|         | Multiple Imputation   | -0.0112 | 0.0024    | 0.6302 | 0.0009     | 0.938    | 0.007    |