Published in final edited form as:

J Appl Stat. 2016 ; 43(15): 2886–2896. doi:10.1080/02664763.2016.1155110.

Multiple Imputation of a Randomly Censored Covariate Improves Logistic Regression Analysis

Folefac D Atem¹, Jing Qian², Jacqueline E Maye³, Keith A Johnson⁴, and Rebecca A Betensky¹

¹ Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA

² Public Health, University of Massachusetts, Amherst, MA, USA

³ Clinical and Health Psychology, University of Florida, Gainesville, FL, USA

⁴ Radiology, Massachusetts General Hospital, Boston, MA, USA

Abstract

Randomly censored covariates arise frequently in epidemiologic studies. The most commonly used methods, including complete case and single imputation or substitution, suffer from inefficiency and bias. They make strong parametric assumptions or they consider limit of detection censoring only. We employ multiple imputation, in conjunction with semi-parametric modeling of the censored covariate, to overcome these shortcomings and to facilitate robust estimation. We develop a multiple imputation approach for randomly censored covariates within the framework of a logistic regression model. We use the non-parametric estimate of the covariate distribution or the semiparametric Cox model estimate in the presence of additional covariates in the model. We evaluate this procedure in simulations, and compare its operating characteristics to those from the complete case analysis and a survival regression approach. We apply the procedures to an Alzheimer's study of the association between amyloid positivity and maternal age of onset of dementia. Multiple imputation achieves lower standard errors and higher power than the complete case approach under heavy and moderate censoring and is comparable under light censoring. The survival regression approach achieves the highest power among all procedures, but does not produce interpretable estimates of association. Multiple imputation offers a favorable alternative to complete case analysis and ad hoc substitution methods in the presence of randomly censored covariates within the framework of logistic regression.

Keywords

Alzheimer's disease; Age of onset; Complete case analysis; Limit of detection; Reverse survival regression

Correspondence to: Rebecca A Betensky.

Conflict of interest:

None declared.

1. Introduction

The risk of Alzheimer's disease (AD) is known to increase dramatically with age¹. Another major risk factor is family history (FH), particularly when it involves the parents²⁻³ and young age of onset⁴⁻⁵. Much of the FH risk has been attributed to the e4 allele of the apolipoprotein E (APOE) genotype⁶. An interaction of APOEe4 effects and gender is widely recognized⁷, and evidence for a maternal transmission factor for AD has been reported⁸⁻⁹. However, controlling for age and female longevity, the data have been inconsistent¹⁰⁻¹¹. AD imaging markers have been investigated to explore the biological basis of these risk factors. AD-like changes in regional brain volume¹²⁻¹³, FDG metabolism¹⁴ and beta-amyloid (A β) deposition, as imaged in vivo with Pittsburgh Compound B (PiB)¹⁵, have been reported in non-demented subjects with maternal FH in excess of what is seen in groups of subjects with no FH or in those with a paternal FH. These changes have been detected even after controlling for APOE e4 carrier status. To further explore this phenomenon, a study was conducted at Massachusetts General Hospital and Brigham and Women's Hospital that investigated the relationship between maternal history of dementia and A β burden in offspring¹⁶.

Beta-amyloid, a continuous measure, is often dichotomized in statistical analyses, to divide subjects into a "high" amyloid group and a "low" amyloid group. This is done because it is not known how to optimally model continuous A β and because there may be a threshold effect, such that the baseline risk of progression of the disease changes from one level to another according to the level of beta-amyloid relative to the threshold. It is also done to standardize different platforms for measurements across studies¹⁷. Thus, we consider that the primary statistical framework to be used to address the scientific question of interest is a logistic regression model for beta-amyloid positivity in the offspring as a function of maternal age of dementia onset, offspring age, gender, education, and Clinical Dementia Rating (CDR) global score. However, as not all mothers experienced onset of dementia at the time of the offspring's interview, their ages at onset are right censored by their ages at the offspring interview, or their ages at death, if this occurred prior to the interview. This introduces the analytical challenge of how to handle randomly right censored covariates in a logistic regression model.

The most commonly used approach for handling censored covariates is the complete case method, which uses only subjects for whom the covariate is not censored. This approach leads to consistent estimation and valid hypothesis testing as long as the censoring distribution is independent of the outcome, given the true value of the covariate (and other covariates in the model). However, it suffers from inefficiency in the presence of moderate to heavy censoring, which we have in our Alzheimer's study. *Ad hoc* substitution methods have been used for limit of detection censoring in the context of linear regression¹⁸. Under limit of detection censoring, all measurements are subject to the same, fixed censoring value. While there are several papers that treat censored covariates in the linear regression framework, we are aware of only two papers that specifically addressed the problem of limit of detection censoring of covariates in logistic regression models^{19,20}. The paper by Cole et al¹⁹ assumed a parametric distribution for the covariate and proposed a maximum likelihood approach for estimation of the regression coefficient of interest. While these authors did not

consider the case of random censoring, their approach could be extended to that setting. The paper by Schisterman et al.²⁰ primarily focused on substitution methods for censored covariates in linear regression models, but also mentioned the computational difficulties that arise for logistic regression models. These are due to the requirement to solve two non-linear equations and the existence of a solution only in rare cases and under strong assumptions. There is one paper²¹ that presents methods for generalized linear regression in the presence of randomly censored covariates, which, of course, includes logistic regression. This paper proposed a quasi-score, estimating equations approach for estimating the regression coefficient for a censored covariate. This weighted least squares regression is equivalent to mean imputation for simple linear regression. This is motivated by computational difficulties that arise when distributions other than the normal are assumed for the censored covariate. The approach assumes a parametric distribution for the censored covariate, such as a generalized gamma distribution. While this offers a robust alternative to maximum likelihood and multiple imputation as it does not require a distributional assumption for the dependent variable, this is not relevant for binary data such as we consider in this paper. A limitation of this paper is its reliance on parametric specification of the distribution of the censored covariate. This might potentially be relaxed, though the authors note that computational and numerical difficulties arise when the distribution of the covariate is not selected as Weibull or generalized gamma, though even the generalized gamma distribution presented numerical difficulties in simulations with small sample sizes²¹.

Our goal for the analysis of the Alzheimer's study is to minimize the assumptions that we make about the data, while optimizing the available information. Thus, we propose a multiple imputation method that is based on the assumed logistic regression model, in conjunction with a nonparametric or semi-parametric estimator for the distribution of the covariate. This improves upon a previous approach that requires a parametric assumption for the distribution of the censored covariate²¹. If there are no other covariates in the model, then the nonparametric Kaplan Meier estimator is used, while if there are other covariates, the Cox model based estimator of adjusted survival is used. Our procedure is computationally tractable and is feasible in relative small samples. We successfully implemented this approach in the context of a linear regression model with a randomly censored covariate²² and showed it to be more efficient than the complete case analysis under moderate censoring. In the absence of parametric assumptions on the distribution of the censored covariate, this approach is not applicable to limit of detection censoring because the nonparametric baseline hazard estimators cannot resolve limit of detection censoring, as there is no information about the covariate beyond the limit of detection. Use of multiple imputation in this setting effectively is assuming a parametric distribution for the covariate that is not based on any data and could therefore lead to bias. Our multiple imputation approach could be taken for limit of detection censoring of a covariate if a parametric distribution were assumed for the covariate.

We compare the proposed multiple imputation approach with the complete case analysis in simulations with light, moderate and heavy censoring. If a test of association between the binary outcome and the covariate is of interest, rather than an estimate of the log odds ratio, then a *reverse survival regression* approach is applicable²². We evaluate the testing properties of this approach, as well, in the simulations. We apply these methods to the

Alzheimer's study of the association between beta-amyloid positivity in offspring and maternal age of onset of dementia.

2. Methods

2.1 Study population

One hundred and forty seven participants were enrolled in the Alzheimer's study, which was approved by the Partners Human Research Committee. All participants were evaluated with interviews, cognitive testing and informant interviews, and judged to be either cognitively normal ($N = 104$) with Clinical Dementia Rating [CDR] 0²³ or mildly impaired ($N = 43$; CDR 0.5). None of the participants had any neurological or medical illness or any history of alcoholism, drug abuse, or head trauma. All scored 11 or lower on the Geriatric Depression Scale [GDS]²², 22 or higher on the Mini Mental State Examination [MMSE]²⁵ and 93 or higher on the American National Adult Reading Test [AMNART]²⁶. A parental history questionnaire yielded information about 279 biological parents of participants. As in previous analyses¹⁷, we used a threshold of 1.2 to dichotomize PiB measurements of beta-amyloid into low and high subgroups. Maternal age at onset of dementia was censored for 70% of participants.

2.2 Statistical methods

We assume the logistic regression model,

$$\Pr(Y=1|X, Z) = \Psi(\alpha_0 + \alpha_1 X + \alpha_2 Z), \quad (1)$$

where Y denotes the binary outcome of interest, X denotes the covariate of interest that is potentially right censored, Z denotes other covariates, and $\Psi(t) = (1 + e^{-t})^{-1}$. X is observed only if $X < C$, where C is the random censoring variable, which we assume is independent of X . In our example, $Y=1$ for amyloid positive individuals and $Y=0$ for those who are amyloid negative, X denotes maternal age at dementia onset, C denotes maternal age at death or at maternal age on the date of the interview of the offspring and Z denotes other covariates. The observed data for each subject are $\{Y, T, Z, D\}$, where $T = \min(X, C)$ and $D = 1$ if $X > C$ and $D = 0$ if $X < C$.

Under the assumption that C is independent of Y given X and Z , it follows from applications of Bayes' theorem that $\Pr(Y=1|X, Z, X < C) = \Pr(Y=1|X, Z)$. This justifies the complete case analysis, which analyzes only those observations with $X < C$. In our example, this translates into independence between maternal age last known to be dementia-free --if the mother did not have dementia onset by the time of offspring's study entry-- and amyloid positivity, given maternal age at dementia onset. While clearly both maternal ages are not observable, as event time and censoring time never are both observable, this seems like a reasonable assumption for this study, which did not recruit subjects on the basis of maternal dementia status.

Although the complete case approach has the advantage of simplicity, it can be inefficient in the presence of moderate censoring²². Thus, we consider the alternative approach of

multiple imputation. Multiple imputation is a principled, likelihood-based method for handling missing data. This means that it is based on models for the data and the methods of estimation and inference are based on formal statistical principles. As a principled approach, and being likelihood based, it leads to efficient and unbiased estimation under certain assumptions on the missingness^{27,28}. The alternative complete case and substitution and single imputation approaches that are used to handle censored covariates do not share these desirable properties. The general multiple imputation scheme consists of three steps: imputation, completed data analysis, and pooling. The imputation step involves first drawing the parameters of the posterior distribution of X given the observed data from their distribution²⁹, and then drawing M sets of imputed values for the missing data from their posterior distribution given the observed data. The first part of this imputation step is necessary to account for the fact that the parameters are unknown and their estimates have inherent variability. In our setting of censored covariates, the observed data consist of $(Y, Z, T=\min(X, C), D=I(C < X))$ and we must condition on the fact that $X > C$ (with C observed) for censored observations. The completed data analysis involves performing the desired analysis on the M completed data sets. The pooling step combines the estimates from the M analyses into a single multiple imputation estimate and likewise combines the estimates of variability from the M analyses, along with the between imputation variability, into a single multiple imputation variance estimate.

In particular, our multiple imputation algorithm is based on full likelihood specification of the observed data and derived conditional (predictive) distributions of X given the observed data, and proceeds in the following steps:

- I) Sample with replacement from the original data.
- II) Fit model (1) using the uncensored observations to effectively sample from the distribution of the coefficients (a_0, a_1, a_2) and obtain estimates $(\hat{\alpha}_0^c, \hat{\alpha}_1^c, \hat{\alpha}_2^c)$.
- III) Fit a Cox model to the sampled data for X given Z to estimate β and $f_{\beta}(x/z)$, the Cox model based estimate of the density of X given Z , with corresponding survivor function, $S_{\beta}(x/z)$.
- IV) Generate X from its predictive density, which we denote as $P(X \in dx | C = c, X > c, Y = y, Z = z)$, and which includes conditioning on Y and is equal to:

$$P(X \in dx | C = c, X > c, Z = z, Y = y) = \frac{\left(\frac{\exp[y(\hat{\alpha}_1^c x + \hat{\alpha}_2^c z)]}{1 + \exp(\hat{\alpha}_1^c x + \hat{\alpha}_2^c z)} \right) \frac{f_{\beta}(x; z)}{S_{\beta}(c; z)} dx}{\int_c \left(\frac{\exp[y(\hat{\alpha}_1^c v + \hat{\alpha}_2^c z)]}{1 + \exp(\hat{\alpha}_1^c v + \hat{\alpha}_2^c z)} \right) \frac{f_{\beta}(v; z)}{S_{\beta}(c; z)} dv}$$

and thus

$$P(X > x | C=c, X > c, Z=z, Y=y) = \frac{\int_x \left(\frac{\exp[y(\hat{\alpha}_1^c v + \hat{\alpha}_2^c z)]}{1 + \exp(\hat{\alpha}_1^c v + \hat{\alpha}_2^c z)} \right) \frac{f_{\beta}(v; z)}{S_{\beta}(c; z)} dv}{\int_c \left(\frac{\exp[y(\hat{\alpha}_1^c v + \hat{\alpha}_2^c z)]}{1 + \exp(\hat{\alpha}_1^c v + \hat{\alpha}_2^c z)} \right) \frac{f_{\beta}(v; z)}{S_{\beta}(c; z)} dv}.$$

In order to avoid estimating the density function, f , we apply integration by parts to the above equation. This serves to replace f with the survivor function, S , in transformed integrals. The survivor function is readily estimable as the Kaplan Meier or as the baseline survivor function from the Cox model. The expression is evaluated numerically at each censored observation and set to a uniform random variate to solve for x , via the inverse cumulative distribution function method for random variable generation³⁰

- V) Fit a logistic regression model for Y given the imputed covariate, X^m , and additional covariates, Z , and estimate parameters of the $(\hat{\alpha}_0^m, \hat{\alpha}_1^m, \hat{\alpha}_2^m)$, where the superscript m labels the estimates from the m th imputation.
- VI) Repeat steps I-V M times.
- VII) Obtain the multiple imputation estimate of α_1 and its standard error³¹ as

$$\hat{\alpha}_1 = \sum \hat{\alpha}_{1m} / M \text{ and}$$

$$Var(\hat{\alpha}_1) = \sum_{m=1}^M Var(\hat{\alpha}_{1m}) / M + (1 + 1/M) \sum_{m=1}^M (\hat{\alpha}_{1m} - \hat{\alpha}_1) / M - 1.$$

Estimation of α_1 and testing whether it is equal to zero is the natural test for association between Y and X , controlling for Z . This is accomplished through the multiple imputation procedure that is outlined above. As an alternative approach to the test for association between Y and X controlling for Z , we use the Cox proportional hazards model

$h(x|y, z) = h_0(x) \exp(\hat{\alpha}_1 y + \hat{\alpha}_2 z)$, where $h(x|y, z)$ is the hazard function for X given Y and Z and $h_0(x)$ is the baseline hazard function for X . This *reverse survival regression* approach has the advantage of automatically and naturally handling the censored X , as it is the outcome, rather than covariate, in the Cox model framework. Under the assumptions of model (1) and independence of C and X given Y and Z , it is straightforward to justify that the test of $H_0: \hat{\alpha}_1 = 0$ based on this model that reverses the natural roles of Y and X yields a valid test for $H_0: \alpha_1 = 0$, where α_1 is the coefficient of covariate X in model (1). The same arguments that we provided about the validity of this test for the linear regression model²² hold in this setting, as well. However, the parameter from the Cox model, $\hat{\alpha}_1$, does not have a meaningful interpretation, and so it is important to note that this simple approach is only applicable for hypothesis testing, and not for estimation.

2.3 Monte Carlo simulation

We evaluated the operating characteristics of the multiple imputation procedure, as well as the complete case analysis, in simulation studies. We additionally applied the reverse survival regression approach to evaluate the type I error and power of the associated Wald

test. As a benchmark, we also analyzed the full data sets using the true values of X , prior to censoring by C .

We assess the effects of sample size and level of censoring. We generated $Z \sim \text{Bernoulli}(0.5)$, $X \sim \text{Weibull}(1, 1/3)$ and $C \sim \text{Weibull}(1, q)$, with $q=1.33$ to obtain light censoring (10-20%), $q=0.40$ to obtain moderate censoring (40%), and $q=0.20$ to obtain heavy censoring (70%) to match the level of censoring in our Alzheimer's study example. We considered sample sizes of $N=150, 200, 500$ and 2000 to obtain a range of values for power, and in the first case ($N=150$), to match the sample size in the Alzheimer's study. We generated Y as Bernoulli, with probability calculated based on the logistic regression model (1), with $(\alpha_0, \alpha_1, \alpha_2) = (-0.75, 1.0, -0.50)$, $(-0.75, 0.60, -0.50)$ and $(-0.75, 0.00, -0.50)$, with the latter selected to match the $\alpha_1 \sim 0$ in the Alzheimer's study. We generated 5000 replications to assess type I error and 1000 replications for power, bias, and standard error estimation.

3. Results

3.1 Simulation results

Table 1a lists the results of the simulations for light censoring (10-20%), Table 1b for moderate censoring (40%), and Table 1c for heavy censoring (70%). The standard errors listed for the multiple imputation approach are averages of the square root of the multiple imputation variances. Under 10-20% censoring (Table 1a), the multiple imputation is comparable to the complete case analysis with regard to bias and exhibits an approximately 30% decrease in standard errors, even for $N=150$. This is true for all three values of α_1 , with perhaps slightly higher absolute bias for α_1 for $N=150$ and 200 , which is on the order of $<5\%$ relative bias. The power of the multiple imputation analysis is comparable to the complete case analysis for $N=150$ and 200 , and may be comparable or slightly larger for $N=500$ and 2000 . As expected, both procedures have diminished power relative to that achievable in the absence of censoring, though these converge with increasing N . The type I errors are very close to the nominal 0.05 level for every N . Under 40% censoring (Table 1b), the advantages of multiple imputation relative to the complete case analysis are more evident than in the light censoring case, and lower standard errors and higher power are obtained for all N . The bias is comparable or smaller, except for $N=150$, for which the bias when $\alpha_1=1.0$ is 0.071 for multiple imputation and 0.048 for complete case analysis, and when $\alpha_1=0.6$ is 0.089 for multiple imputation and 0.064 for complete case analysis. Note that relative to the magnitudes of α_1 , these are quite small. Even under 70% censoring (Table 1c) that mimics the Alzheimer's study, there are clear and consistent advantages to multiple imputation over complete case analysis for all N and α_1 . For all scenarios, the reverse survival regression approach attains higher power than the multiple imputation and complete case approaches, though this is at the expense of providing an estimate of the log odds ratio parameter of interest.

3.2 Alzheimer's study results

Of the 147 participants in the Alzheimer's study, six were excluded because they lacked parents' age at censoring or onset of dementia. Table 2 lists descriptive statistics for the covariates of interest that were included in the logistic regression and Cox regression

models. Table 3 provides the estimates, standard errors, and p-values for the log odds ratio association parameter of interest, α_1 . As expected, the standard error for the complete case analysis (0.041) is larger than that for the multiple imputation analysis (0.017). We find no significant association between beta-amyloid positivity and maternal age of onset of dementia in this study. This result is consistent across the complete case analysis, the multiple imputation analysis, and the reverse survival regression.

4. Discussion

We have developed a multiple imputation procedure for handling randomly right censored covariates in a logistic regression model. This advances the literature in this area by allowing for random censoring and by treating the covariates nonparametrically or semi-parametrically. The multiple imputation estimates have lower standard errors than those from the complete case analysis, at the expense of possible very slight increased absolute bias, but very low relative bias. They exhibit increased power relative to complete case. The reverse survival method consistently has higher power than both the multiple imputation and complete case based tests. However, the reverse survival method does not produce interpretable estimates of the log odds ratio of interest, and thus multiple imputation is preferable when estimation is of interest.

Acknowledgments

Funding

This work was supported by National Institutes of Health grants 5T32NS048005, 5P50AG005134, 5P01AG036694; Harvard NeuroDiscovery Center

References

1. Mayeux R. Epidemiology of neurodegeneration. *Annu Rev Neurosci.* 2003; 26:81104.
2. Jarvik LF, Blazer D. Children of Alzheimer patients: an overview. *Journal of Geriatric Psychiatry and Neurology.* 2005; 18:181186.
3. Jarvik L, LaRue A, Blacker D, Gatz M, Kawas C, McArdle JJ, et al. Children of persons with Alzheimer disease: what does the future hold? *Alzheimer disease and associated disorders.* 2008; 22:620.
4. Silverman JM, Smith CJ, Marin DB, Mohs RC, Propper CB. Familial patterns of risk in very late-onset Alzheimer disease. *Arch Gen Psychiatry.* 2003; 2003; 60:190197.
5. Silverman JM, Ciresi G, Smith CJ, Marin DB, Schnaider-Beeri M. Variability of familial risk of Alzheimer disease across the late life span. *Arch Gen Psychiatry.* 2005; 62:565573.
6. Mayeux R. Clinical practice. Early Alzheimer's disease. *N Engl J Med.* 2010; 362:21942201.
7. Miech RA, Breitner JCS, Zandi PP, Khachaturian AS, Anthony JC, Mayer L. Incidence of AD may decline in the early 90s for men, later for women: The Cache County study. *Neurology.* 2002; 58:209218.
8. Duara R, Lopex-Alberola RF, Barker WW, Loewenstein DA, Zatinsky M, Eisdorfer CE, Weinberg GB. A comparison of familial and sporadic Alzheimers disease. *Neurology.* 1993; 43:1377–1384. [PubMed: 8327141]
9. Edland SD, Silverman JM, Peskind ER, Tsuang D, Wijsman E, Morris JC. Increased risk of dementia in mothers of Alzheimer's disease cases Evidence for maternal inheritance. *Neurology.* 1996; 47:254–256. [PubMed: 8710088]
10. Heggeli KA, Crook J, Thomas C, Graff-Radford N. Maternal Transmission of Alzheimer Disease. *Alzheimer Disease and Associated Disorders.* 2012; 26:364–366. [PubMed: 22273801]

11. Ehrenkrantz D, Silverman JM, Smith CJ, Birstein S, Marin D, Mohs RC, Davis KL. Genetic epidemiological study of maternal and paternal transmission of Alzheimer's disease. *Am J Med Genet.* 1999; 88:378382.
12. Honea RA, Swerdlow RH, Vidoni ED, Goodwin J, Burns JM. Reduced gray matter volume in normal adults with a maternal family history of Alzheimer disease. *Neurology.* Jan 12. 2010; 74(2):113–20. [PubMed: 20065246]
13. Berti V, Mosconi L, Glodzik L, Murray J, De Santi S, Pupi A, Tsui W, DeLeon MJ. Structural brain changes in normal individuals with a maternal history of Alzheimer's. *Neurobiol Aging.* Dec; 2011 32(12):2325, e17–26.
14. Mosconi L, Brys M, Switalski R, Mistur R, Glodzik L, Pirraglia E, Tsui W, De Santi S, de Leon MJ. Maternal family history of Alzheimer's disease predisposes to reduced brain glucose metabolism. *Proc Natl Acad Sci USA.* 2007; 104:19067–19072. [PubMed: 18003925]
15. Mosconi L, Rinne JO, Tsui WH, Berti V, Li Y, Wang, et al. Increased fibrillar amyloid- β burden in normal individuals with a family history of late-onset Alzheimer's. *Proceedings of the National Academy of Sciences.* 2010; 107:5949–5954.
16. Maye JE, Gidyczin C, Pepin LE, Becker AJ, Blacker D, Rentz D, Locascio J, Sperling RA, Johnson K. Maternal dementia age of onset in relation to amyloid burden in non-demented offspring. 6th Annual Human Amyloid Imaging Meeting; Miami. January 2012; 2012.
17. Mormino EC, Betensky RA, Hedden T, Schultz AP, Ward A, Huijbers W, Rentz DM, Johnson KA, Sperling RA. Alzheimer's Disease Neuroimaging Initiative; Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing; Harvard Aging Brain Study. Amyloid and APOE ϵ 4 interact to influence short-term decline in preclinical Alzheimer disease. *Neurology.* May 20. 2014; 82(20):1760–7. [PubMed: 24748674]
18. Hornung RW, Reed LD. Estimation of average concentration in the presence of nondetectable values. *Appl Occup Environ Hyg.* 1990; 5(1):46–51.
19. Cole SR, Chu H, Nie L, Schisterman E. Estimating the odds ratio when exposure has a limit of detection. *International Journal of Epidemiology.* 2009; 38:1674–1680. [PubMed: 19667054]
20. Schisterman EF, Vexler A, Whitcomb BW, Liu A. The limitations due to exposure detection limits for regression models. *Am J Epidemiol.* 2006; 163:374–383. [PubMed: 16394206]
21. Tsimikas JV, Bantis LE, Georgiou SD. Inference in generalized linear regression models with a censored covariate. *Computational Statistics and Data Analysis.* 2011; 56:1854–1868.
22. Atem FD, Qian J, Maye JE, Johnson KA, Betensky RA. Linear Regression with a Randomly Censored Covariate: Application to an Alzheimer's Study. 2015 Submitted.
23. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology.* Nov.1993 43(11):24122414.
24. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO. Development and validation of a geriatric depression screening scale: a preliminary report. 1982; 17(1):3749.
25. Folstein MF, Folstein SE, McHugh PR. Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *Nov.1975* 12(3):189198.
26. Ryan JJ, Paolo AM. A screening procedure for estimating premorbid intelligence in the elderly. *The Clinical Neuropsychologist.* 1992; 6(1):53–62.
27. Kenward MG, Carpenter J. Multiple imputation: current perspectives. *Statistical Methods in Medical Research.* 2007; 16:199–218. [PubMed: 17621468]
28. Rubin, DB. Multiple imputation for nonresponse in survey. John Wiley; New York: 1987.
29. D'Angelo G, Weissfeld L, Chu L. An index approach for the cox model with left censored covariates. *Statistics in Medicine.* 2008; 27:4502–4514. [PubMed: 18407573]
30. Thisted, RA. Elements of statistical computing: numerical computation. Vol. 1. CRC Press; 1988.
31. Schafer JL. Multiple imputation: a primer. *Statistical methods in medical research.* 1999; 8.1:3–15. [PubMed: 10347857]

Table 1a

Estimation and testing of α_1 under light (10–20%) censoring (1000 replicates; 5000 for type I error)

N	Method	$\alpha_1=1.0$				$\alpha_1=0.6$				$\alpha_1=0.0$			
		bias	SE	MSE	power	bias	SE	MSE	power	bias	SE	MSE	type I error
150	No Censoring	0.015	0.551	0.303	0.494	-0.016	0.548	0.300	0.228	-0.057	0.546	0.301	0.043
	Complete Case	0.017	0.774	0.599	0.303	-0.041	0.781	0.611	0.132	-0.063	0.749	0.565	0.040
	Multiple Imputation	0.046	0.640	0.412	0.321	0.052	0.647	0.422	0.128	0.054	0.677	0.461	0.041
	Reverse Survival				0.393				0.166				0.052
200	No Censoring	0.013	0.472	0.223	0.629	0.010	0.456	0.208	0.276	-0.052	0.491	0.244	0.051
	Complete Case	0.015	0.661	0.437	0.374	0.017	0.641	0.411	0.176	-0.060	0.716	0.516	0.050
	Multiple Imputation	0.040	0.542	0.295	0.391	0.020	0.598	0.358	0.166	-0.055	0.577	0.336	0.047
	Reverse Survival				0.490				0.234				0.055
500	No Censoring	-0.002	0.291	0.085	0.933	-0.005	0.281	<0.001	0.579	-0.025	0.311	0.097	0.050
	Complete Case	-0.007	0.405	0.164	0.677	-0.005	0.391	0.153	0.317	-0.021	0.434	0.189	0.053
	Multiple Imputation	0.006	0.326	0.106	0.667	-0.015	0.295	0.087	0.364	-0.010	0.351	0.123	0.046
	Reverse Survival				0.872				0.486				0.052
2000	No Censoring	0.000	0.144	0.011	1.000	0.001	0.139	0.019	0.985	<0.001	0.153	0.023	0.047
	Complete Case	0.006	0.199	0.040	0.990	0.002	0.193	0.037	0.873	<0.001	0.214	0.046	0.054
	Multiple Imputation	-0.001	0.159	0.025	1.000	-0.001	0.150	0.023	0.902	<0.001	0.171	0.029	0.045
	Reverse Survival				1.000				0.970				0.048

Table 1b

Estimation and testing of α_1 under moderate (40%) censoring (1000 replicates; 5000 for type I error)

N	Method	$\alpha_1=1.0$				$\alpha_1=0.6$				$\alpha_1=0.0$				type I error
		bias	SE	MSE	power	bias	SE	MSE	power	bias	SE	MSE	power	
150	No Censoring	0.015	0.551	0.303	0.494	-0.016	0.548	0.301	0.228	-0.057	0.546	0.301	0.228	0.043
	Complete Case	0.048	1.426	2.034	0.109	0.064	1.474	2.177	0.061	-0.099	1.554	2.425	0.040	0.040
	Multiple Imputation	0.071	0.839	0.709	0.123	0.089	0.857	0.743	0.068	-0.056	0.909	0.829	0.059	0.059
	Reverse Survival				0.267				0.124					0.046
200	No Censoring	0.013	0.472	0.223	0.629	0.010	0.456	0.208	0.276	-0.052	0.491	0.244	0.051	0.051
	Complete Case	0.021	1.213	1.471	0.141	0.023	1.165	1.357	0.063	0.093	1.317	1.743	0.049	0.049
	Multiple Imputation	0.036	0.705	0.498	0.189	-0.025	0.827	0.684	0.062	-0.032	0.755	0.571	0.046	0.046
	Reverse Survival				0.362				0.163					0.057
500	No Censoring	-0.002	0.291	0.085	0.933	-0.005	0.281	0.079	0.579	-0.025	0.311	0.097	0.050	0.050
	Complete Case	0.008	0.723	0.523	0.277	0.033	0.700	0.491	0.139	-0.041	0.749	0.563	0.046	0.046
	Multiple Imputation	0.017	0.414	0.172	0.577	-0.020	0.405	0.164	0.246	0.030	0.441	0.195	0.045	0.045
	Reverse Survival				0.725				0.353					0.049
2000	No Censoring	0.000	0.144	0.011	1.000	0.001	0.139	0.019	0.985	<0.001	0.153	0.023	0.047	0.047
	Complete Case	0.007	0.354	0.125	0.800	0.011	0.343	0.117	0.462	-0.017	0.380	0.145	0.048	0.048
	Multiple Imputation	0.003	0.197	0.039	0.990	0.008	0.174	0.030	0.866	-0.002	0.211	0.045	0.052	0.052
	Reverse Survival				1.000				0.910					0.051

Table 1c

Estimation and testing of α_1 under heavy (70%) censoring (1000 replicates; 5000 for type I error)

N	Method	$\alpha_1=1.0$				$\alpha_1=0.6$				$\alpha_1=0.0$			
		bias	SE	MSE	power	bias	SE	MSE	power	bias	SE	MSE	type I error
150	No Censoring	0.015	0.551	0.303	0.494	-0.016	0.548	0.301	0.228	-0.057	0.546	0.301	0.043
	Complete Case	-0.072	2.752	7.577	0.065	-0.224	2.842	8.126	0.055	-0.242	2.706	7.381	0.031
	Multiple Imputation	0.091	1.212	1.477	0.065	0.133	1.229	1.527	0.056	0.061	1.255	1.579	0.048
	Reverse Survival				0.168				0.094				0.053
200	No Censoring	0.013	0.472	0.223	0.629	0.010	0.456	0.208	0.276	-0.052	0.491	0.244	0.051
	Complete Case	0.067	2.246	5.048	0.068	0.074	2.279	5.197	0.059	-0.350	2.405	5.905	0.031
	Multiple Imputation	0.089	0.994	0.996	0.102	-0.077	1.009	1.023	0.059	-0.055	1.038	1.080	0.059
	Reverse Survival				0.238				0.110				0.053
500	No Censoring	-0.002	0.291	0.085	0.933	-0.005	0.281	0.079	0.579	-0.025	0.311	0.097	0.050
	Complete Case	-0.028	1.307	1.710	0.126	-0.067	1.342	1.081	0.069	-0.157	1.412	2.018	0.047
	Multiple Imputation	0.071	0.567	0.327	0.318	0.073	0.576	0.337	0.071	0.031	0.594	0.354	0.057
	Reverse Survival				0.580				0.251				0.056
2000	No Censoring	0.000	0.144	0.011	1.000	0.001	0.139	0.019	0.985	<0.001	0.153	0.023	0.047
	Complete Case	0.007	0.630	0.397	0.370	-0.013	0.644	0.415	0.182	-0.059	0.671	0.454	0.047
	Multiple Imputation	0.009	0.259	0.067	0.929	0.035	0.263	0.070	0.488	0.007	0.273	0.075	0.056
	Reverse Survival				0.988				0.717				0.053

Table 2

Alzheimer's study: descriptive statistics

	Mean (range) for continuous variables;
Variable	Frequency (percentage) for categorical variables
PiB (beta-amyloid)	1.2 (0.98-2.04)
PIB dichotomous	<1.2(68.8%), 1.2 (31.3%)
Education (years)	16.1(9-21)
CDR	0(71.6%), 1(28.4%)
Gender (female=1)	0(46.1%), 1(53.9%)
Offspring age	81(26.2%), <81(73.8%)

Table 3

Application to study of beta-amyloid and maternal age of onset of dementia: estimation and testing of α_1

Method	Estimate	SE	P-value	% deleted
Complete Case	−0.052	0.041	0.252	70.20%
Multiple Imputation	−0.008	0.017	0.504	
Reverse Survival			0.296	