Published in final edited form as:
J Stat Theory Pract. 2013 January 1; 7(2): 381–400. doi:10.1080/15598608.2013.772830.

Bias Correction Methods for Misclassified Covariates in the Cox Model: comparison offive correction methods by simulation and data analysis

Heejung Bang[1,*], **Ya-Lin Chiu**[2], **Jay S. Kaufman**[3], **Mehul D. Patel**[4], **Gerardo Heiss**[4], and **Kathryn M. Rose**[5]

[1]Division of Biostatistics, Department of Public Health Sciences, University ofCalifornia, Davis, CA, USA

[2]Division of Biostatistics and Epidemiology, Department of Public Health, WeillCornell Medical College, New York, NY, USA

[3]Department of Epidemiology, Biostatistics and Occupational Health, McGillUniversity, Montreal, Quebec, Canada

[4]Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA

[5]SRA International, Inc., Durham, NC, USA

Abstract

Measurement error/misclassification is commonplace in research when variable(s) can notbe measured accurately. A number of statistical methods have been developed to tackle this problemin a variety of settings and contexts. However, relatively few methods are available to handlemisclassified categorical exposure variable(s) in the Cox proportional hazards regression model. Inthis paper, we aim to review and compare different methods to handle this problem - naïvemethods, regression calibration, pooled estimation, multiple imputation, corrected score estimation,and MC-SIMEX - by simulation. These methods are also applied to a life course study with recalleddata and historical records. In practice, the issue of measurement error/ misclassification should beaccounted for in design and analysis, whenever possible. Also, in the analysis, it could be moreideal to implement more than one correction method for estimation and inference, with properunderstanding of underlying assumptions.

Keywords

ARIC; Childhood SES; Cox proportional hazards regression; Measurement error; Misclassification; Recalled error

1. Introduction

Measurement error (ME) is common in biomedical and epidemiologic research. When anexposure variable (or covariate) is analyzed as a categorical variable, the ME is generally referredto as ‘misclassification’. Currently, a number of methods have been developed tohandle different types of MEs, study designs and statistical or data settings. Some of these methodsdeveloped from fundamentally different formulations or paradigms, while others are major or minorextensions of extant methods. Most currently available methods are suited for

[*]Corresponding author: Heejung Bang PhD, One Shields Avenue, Med Sci 1-C,Davis, CA 95616-8638. hbang@ucdavis.edu, Phone: (919) 423-2271 Fax: (530)752-3239.

handling continuouscovariate(s) in generalized linear models (GLMs) (e.g., linear or logistic regression) (Freedman et al., 2008; Messer andNatarajan, 2008), while there have been fewer developments for applications with categoricalcovariate and/or censored outcome data. In this paper, we review and compare available methods bysimulation and data analysis that could handle misclassified binary exposure variable(s) in the Coxproportional hazards regression model (Cox, 1972). Weselected five fundamentally different but practical methods - 1) regression calibration; 2) pooledestimation; 3) multiple imputation; 4) corrected score estimation; and 5) MC-SIMEX, and comparedthem to naïve methods that do not account for misclassification properly.

To our knowledge, no prior publication has compared these methods altogether in anycontext. Based on our review, we found that the most common practice in statistical as well asapplied research is to implement only one error correction method and to contrast results before andafter correction. Since the ME correction methods heavily rely on assumptions (some of which are notempirically verifiable), it may be more reasonable to explore/implement different methods ratherthan to use a single method, often chosen by computational convenience or users'familiarity, preference or tradition.

The paper is organized as follows. In Section 2, we briefly review statistical methods.We summarize simulation results in Section 3 and data analysis in Section 4. Section 5 providesdiscussion and conclusion.

2. Data Settings and Statistical Methods: a Review

We adopt a *standard* survival analysis setup, denoting the survival timeby T_i^0 and the time of right censoring by C_i forith individual (i=1,…,n) and the observed data are the minimum of thesetwo times, $T_i = \min(T_i^0, C_i)$ and the event indicator $\Delta_i = I(T_i^0 \leq C_i)$. Survival and censoring processes are conditionally independent giventhe covariate process as in classical survival analysis settings (Kalbfleisch and Prentice, 2002).

The true covariate or gold standard measure is denoted by X. Given that it is oftendifficult or expensive to measure X accurately, we may measure W as a proxy. For example, X is thetrue vitamin D intake and W is a proxy for X, based on assessment of vitamin D intake through a foodfrequency questionnaire or food diary. In the motivating example that we will analyze later, X isfather's occupation during childhood and W is recalled data during adulthood.

Let us suppose that X and W are binary and that the relationship of X and W or themisclassification pattern can be characterized by sensitivity (Se) and specificity (Sp):

$$P(W=1|X=1)=\mathrm{Se} \text{ and } P(W=0|X=0)=\mathrm{Sp}.$$

We assume the misclassification pattern is 'non-differential' forsurvivors and non-survivors. This assumption tends to hold in prospective cohort studies, comparedto case-control studies, where survival analysis is typically conducted (Carroll et al., 1995).

We also use a standard ME setting, where a set of observations$\{T_i, \Delta_i, W_i\}$are available in the full sample (fori=1,…,n), while X_i is additionally available for asubsample (i.e., a validation sample). In this manuscript, we assume a simple setting with thefollowing conditions: 1) there is one error-prone covariate; 2) the covariate is time-invariant; and3) internal validation sample is available, for simpler presentation and comparison. Extensions

tomore advanced or general settings such as those with time-dependent covariate or multiple covariateswith or without ME/misclassification could be made for some methods.

We work under the Cox proportional hazards model with the hazard function of

$$\lambda(t|X)=\lambda_0(t)\exp(\beta X) \quad (2.1)$$

where $\lambda_0(t)$ is aunspecified baseline hazard function and β is an unknown regressionparameter of interest. Our goal is to estimate the point and interval estimates of trueβ with minimal bias.

2.1. Regression calibration

Regression calibration (RC) is a standard method for correcting for bias due to ME(Armstrong, 1985; Carroll andStefanski, 1990; Fuller, 1987; Gleser, 1990; Rosner et al., 1989). RCis a simple and general method, which can be potentially applicable to any regression model. Thebasic idea behind RC is that one replaces X by the regression of X given W (or given W and othercomplete covariates) as an approximation and then performs a standard analysis. Thus, this methodrelies on the assumption that this approximation is sufficiently accurate.

Rosner and colleagues (Rosner et al., 1989)proposed the following simple formulas for the relative risk model with one covariate:

$$\widehat{\beta}_{RC}=\widehat{\beta}_W/\widehat{\gamma}$$

with

$$\widehat{Var}(\widehat{\beta}_{RC})=1/\widehat{\gamma}^2\ \widehat{var}(\widehat{\beta}_W)+\widehat{\beta}_W^2/\widehat{\gamma}^4\ \widehat{var}(\widehat{\gamma})$$

where $\hat{\beta}_W$ is estimated from(2.1) by using W in place of X, and $\hat{\gamma}$ is obtained from fitting the simple linear regression model forX and W:

$$E(X|W)=\alpha+\gamma W$$

under constant variance, $Var(X|W)=\sigma^2$,to the validation sample.

The behavior of bias due to ME in the Cox model has been investigated (Prentice, 1982). Later, It has been noted that the use of Rosner'sformulas can be justified in the Cox model when the following assumptions are met: 1) X iscontinuous; 2) the event is rare; 3) relative risk is small; 3) ME is not severe; 4) ME is additive;and 5) ME is non-differential (Spiegelman, 1997).Additionally, censoring is assumed to be conditionally independent of the true exposure X, given themismeasured exposure W, analogous to the conditional independent censoring assumptions invoked whenstandard survival analysis methods are used with perfectly measured covariate (Kalbfleisch and Prentice, 2002; Spiegelman,1997).

Yet, since RC is easy to understand and convenient to use, possibly the most popularmethod in the ME literature, it is commonly considered for handling discrete covariates ornon-normal data as well (Cole et al., 2006; Dalen et al., 2006).

2.2. Pooled estimation

A pooled estimator, which combines the RC estimator and an estimator from the validationdata has been proposed as well (Spiegelman et al., 2001). Thepooled estimator is formulated as:

$$\widehat{\beta}_{pooled}=w_{RC}\widehat{\beta}_{RC}+w_{V}\widehat{\beta}_{V}$$

where $w_{RC}=\widehat{Var}(\hat{\beta}_{RC})^{-1}[\widehat{Var}(\hat{\beta}_{RC})^{-1}+\widehat{Var}(\hat{\beta}_{RC})^{-1}]^{-1}$and $w_{V}=1-w_{RC}\widehat{Var}(\hat{\beta}_{RC})$

and the corresponding asymptotic variance is given as:

$$\widehat{Var}(\widehat{\beta}_{pooled})=[\widehat{Var}(\widehat{\beta}_{RC})^{-1}+\widehat{Var}(\widehat{\beta}_{V})^{-1}]^{-1}$$

where $\hat{\beta}_{RC}$ is the standard RCestimator from Section 2.1 and $\hat{\beta}_{RC}$ is the slopeestimator obtained from the validation data alone from the primary regression model (2.1).

This extension leads to increased efficiency compared to the standard RC estimator whenthe validation sample is large. Selecting an appropriately large validation sample is important inthe context of the Cox model although it is not always feasible or practical.

Regarding censoring mechanism, the same censoring assumption for RC above is assumed inthe main study, while censoring in the internal validation sample would be conditionally independentgiven the true exposure as we just do standard survival data analysis on the true exposure ignoringthe mismeasured exposure entirely in the internal validation sample.

2.3. Multiple imputation

Multiple imputation (MI) was originally developed to solve missing data problems instatistics (Little and Rubin, 2002; Rubin, 1976). Yet, considerable similarities in missing and mismeasured datahave been noted and some methods can handle these two types of incomplete data together. Among anumber of statistical methods for the analysis of missing data, MI is popularly employed, partlybecause the operating mechanism is intuitive (e.g., filling in the missing data by artificial butplausible data multiple times and combining the results) and also because it is flexible and easy toimplement for a variety of statistical models. The use of MI has been suggested as a bias correctionmethod for a binary covariate subject to misclassification in the Cox model (Cole et al., 2006). We recap the general algorithm below, which can bemodified to accommodate different models as needed.

Step 1—Fit a logistic regression model that relates X to W in the validation sample:

$$\text{logit}\,\Pr(X=1|W=w,\Delta=\delta,T=t)=\alpha_0+\alpha_1 w+\alpha_2\delta+\alpha_3 f(t)$$

where f is a function such as identity, log or spline. Then store theresulting parameter estimates (i.e., $\hat{\alpha}_0,\hat{\alpha}_1,\hat{\alpha}_2,\hat{\alpha}_3$) and covariance matrix (say,$\hat{\Sigma}_{w,\delta,t}$).

[Remark: In this regression, one can add the interaction of w and δ orother observed covariates, where the interaction term can partly address differentialmisclassification.]

Step 2—Using the estimated parameters and covariance matrix, draw an estimate of the set offour coefficients for each imputation k (k=1,…,K) from a multivariate normaldistribution with mean vector $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)$ and covariance matrix $\hat{\Sigma}_{w,\delta,t}$.

Step 3—Let $Z_k = X$ whenever X is available (that is, in the validation sample). If not, draw $Z_k \sim Bernoulli(\hat{p}_{k,w,d,t})$, where $\hat{p}_{k,w,\delta,t} = 1/[1+\exp\{-(\hat{\alpha}_{0,k} + \hat{\alpha}_{1,k}\, w + \hat{\alpha}_{2,k}\, \delta + \hat{\alpha}_{3,k} I(t))\}]$ for each k=1, …, K. Now K imputed datasetsare ready.

[Remark: If computing resource and time are not a major issue, we suggest amoderate to large number of imputations (say, K=10–40) as Cole et al. recommendedrather than traditionally recommended number such as 5 in the missing data literature.]

Step 4—Fit K models separately and then combine the results. Explicitly, fit a Cox model $\lambda(t|Z_k) = \lambda_{0,k}(t)\exp(\beta_k Z_k)$ for k=1 to K. Then the final hazardratio and its variance can be estimated by the standard combining schemes in MI:

$$\exp(\bar{\beta}) = \exp\left(\sum_{k=1}^{K} \hat{\beta}_k / K\right)$$

where $\hat{\beta}_k$ is the log hazardratio obtained from the kth imputed dataset in Step 3, and

$$Var(\bar{\beta}) = \sum_{k=1}^{K} \widehat{Var}(\hat{\beta}_k)/K + \sum_{k=1}^{K} (\hat{\beta}_k - \bar{\beta})^2 * (1+1/K)/(K-1)$$

which combines variability within- and between-imputations.

Currently, many standard statistical software packages (e.g., MI and MIANALYZEprocedures with CLASS statement in SAS) provide user-friendly commands for implementing MI. Thereare some conceptual advantages in this method as well: MI uses true exposure whenever it isavailable, and differential ME (for event vs. non-event) are typically better handled by missingdata methods than standard ME methods (Carroll, 2005; White, 2006). Yet, the correct specification of the models iscritical for successful performance of this method, and MI with censored outcomes is more difficultto implement than applications without censored data in general (Qiet al., 2010; Van Buuren et al., 1999; White, 2006).

2.4. Corrected score estimation

The corrected score (CS) estimator was proposed for the Cox model with misclassifieddiscrete covariates (Zucker and Spiegelman, 2008) byextending the original CS techniques (Akazawa et al., 1998;Nakamura, 1990). Under the Cox model in (2.1) in the absence of ME, the partial likelihood score function canbe written as:

$$score(\beta) = \frac{1}{n}\sum_{i=1}^{n} \Delta_i \left(X_i - \frac{1/n \sum_{j=1}^{n} Y_j(t) X_i \exp(\beta X_i)}{1/n \sum_{i=1}^{n} Y_j(t) \exp(\beta X_i)} \right).$$

The basic idea is that all terms that include X (i.e., X_i,exp(βX_i), $X_i$exp(βX_i)) are replaced by observable quantities, and theresulting score is called 'CS function'. For example, unobservedX_i is replaced by *observable* functiong*(W)=B f(W), where B is a function of the misclassification matrixΠ, which consists of Se and Sp, and f is some function. Here, the novel device B is chosento make the key relationship E[g*(W)|X]=g(X) hold. In theabsence of misclassification, this method reduces to the classical Cox partial likelihood method. Asandwich formula and bootstrap are suggested for variance estimation.

With this method, instead of using the individual (raw) data from the validation sample,Se and Sp estimated from this sample are used. This results in some loss of efficiency but couldaccommodate situations where a validation sample is formed using a nonrandom or nonrepresentativesubset of study participants (e.g., those who died). Also, CS is a 'functional'modeling approach unlike RC and MI in the sense that knowledge about the distribution of Xs isavoided. However, when the risk sets get small, say, in the right tail of the time axis, somenumerical problems could occur. Generally, administrative truncation can make risk sets sufficientlylarge enough to resolve this problem, which is not uncommon in survival analysis (Bang, 2005; Huang and Wang, 2001).Notably, this method allows the censoring to depend on X but does not allow it to depend on W,differently from other methods.

2.5. MC-SIMEX

Simulation and extrapolation (SIMEX) is another general method that can deal withadditive ME in continuous variable (Cook and Stefanski,1995). This method consists of 'simulation' and'extrapolation' steps, and is particularly useful for complex models with a simpleME structure. Later, SIMEX has been extended to handle misclassification of categorical variablesand called the method, MC-SIMEX (Kuchenhoff et al.,2006).

The key idea is that SIMEX estimates are obtained by *adding* additionalME to the data like resampling, establishing a trend of ME-induced bias over the variance of theadded ME, and then extrapolating this trend back to the case of no ME.

For a continuous covariate, SIMEX uses the relationship between the size of the ME,denoted by σ_u^2 and the bias in the parameter estimator. We may define a function:

$$\sigma_u^2 \to \beta^*(\sigma_u^2) := f(\sigma_u^2)$$

where β^* is the limit towhich the naïve estimator converges as $n \to \infty$,$f(0) = \beta$, the true parameter, and $f(\sigma_u^2)=\widehat{\beta}_W$, the naïve estimator. SIMEX tries to approximate the function$f(\cdot)$ by a parametric approach, for example, via linear, quadratic or logfunction. Then extra ME, $\lambda\sigma_u^2$, is added to W by 'simulation' so that the resulting MEis $(1+\lambda)\sigma_u^2$ and the corresponding estimator is $f((1+\lambda)\sigma_u^2)$. Repeating this simulation step for a fixed grid ofλ will generate the data pairs for (λ, $f(\sigma_u^2)$) and then we may fit the function $f(\cdot)$, say,by least squares. Finally, we have the SIMEX estimator$\hat{\beta}_{SIMEX} = \hat{f}(0)$ when$\lambda = -1$, that is, the approximated function is'extrapolated' back to the *hypothetical* situation, where there isno ME. A graph is often drawn with parameter estimate for Y-axis and λ forX-axis (say, for $-1 < \lambda < 2$), where the y-value for$\lambda = -1$ and 0 is the SIMEX estimate and naïveestimate, respectively, and $\lambda > 0$ corresponds to simulated situationswith increased ME.

For a binary covariate, the misclassification error can be described by themisclassification matrix Π instead of σ_u^2. Using a similar logic outlined above, the MC-SIMEX estimator can bedefined by a parametric approximation of :

$$\lambda \to \beta^*(\Pi^\lambda) := f(1+\lambda)$$

where π^λ can be expressed as$\pi^\lambda := E\Lambda^\lambda E^{-1}$ via spectraldecomposition, with Λ being the diagonal matrix of eigenvalues and E the correspondingmatrix of eigenvectors. Then by performing a similar simulation step (i.e., generate pseudo data andcompute the naïve estimators for each λ) and extrapolation step(i.e., fit a curve for the relationship of X= λ vs. Y=$f(1 + \lambda)$ and find the Y value that corresponds toX = $\lambda = -1$ as in the SIMEX), the MC-SIMEX estimatoris computed as $\hat{\beta}_{MC-SIMEX} = \hat{f}(0)$.

Three variance estimation methods have been proposed: jackknife, asymptotic andbootstrap (Kuchenhoff et al., 2007; Kuchenhoff et al., 2006). The SIMEX methods rely on simulation andextrapolation functions, based on a premise that the effect of ME on an estimator can be determinedexperimentally via simulation. Thus, they do not necessarily yield a consistent estimator andextrapolation process could be numerically unstable (Lederer andKuchenhoff, 2006). Kuchenhoff et al. (2006) studiedGLMs but the same logic could be extended to survival regression (Slate and Bandyopadhyay, 2009).

3. Simulation

We conducted a simulation study for a simple Cox regression model with one covariate. Weevaluated the performances of two naïve methods (using the observed misclassified covariate,W, and using the true covariate, X, in the validation sample only) and five correction methods(denoted by RC, Pooled, MI, CS, and MC-SIMEX), which were compared to the hypothetical situationwhen X is available for all subjects.

A binary X was generated from a Bernoulli distribution with the prevalence,P(X=1)=0.4 or 0.2. The survival time, T, was generated from a Weibull distributionwith the shape parameter of 2 and the scale parameter of exp(a-log(1.5)*X) that yields thetrue hazard ratio (HR) of 2.25, or equivalently, log (HR)=0.81, where a=1.9 was usedfor common event scenarios and a=1 for rare event scenarios. Censoring time, C, wasgenerated from an exponential distribution with the mean of 1, independently from all othervariables – we will discuss the situation when censoring depends on covariate at the end ofthis section. Then, the follow-up time was defined as the minimum of the survival time and censoringtime. We created misclassified W from X according to Se and Sp parameters (see Table 1 for simulation configurations).

To summarize briefly, we used 40% and 20% for prevalence of the trueexposure, 2000 and 1000 for the sample size, n, of the full sample, approximately 20% and5% for the event rate, and (0.9, 0.7), (0.9, 0.9) and (0.7, 0.9) for (Se, Sp). Out of allpossible combinations, we reported in Table 1 the 10scenarios that were deemed to be most important in practice. We also added one additional simulationscenario that closely characterizes our example (that is, 16% of the exposure prevalence,n=5000, the event rate of 7%, Se=0.55 and Sp=0.80). For allsimulations, 10% subsample was randomly selected for the purpose of validation.

Simulation was repeated 1000 times and results were summarized in terms of 1) mean of(absolute) bias estimates in log(HR); 2) sample standard error (SSE); 3) mean of standard errorestimates (SEE); 4) mean squared error (MSE); and 5) coverage probability (CP). Of note, 20imputation datasets were generated for MI and 100 simulations with quadratic

extrapolation functionwere used for MC-SIMEX. Also, we used the Poisson approximation of the Cox model in theimplementation of MC-SIMEX as the current method and software are not directly applicable to the Coxmodel (Lindsey, 1995; Loomiset al., 2005).

We repeated the same set of simulations with a more modest but protective effect size(HR=0.84) and presented the results in Table 2.[Remarks: We used n≥1000 because ME correction is generally applied to largeepidemiologic studies and statistical power is governed by the number of the events in survivalanalysis. In small or moderate size studies (say, N<500), particularly with rare events, where itis likely that only few people in the validation sample might have the event, the ME correction maynot be feasible or reliable for the Cox model. Also, we chose 10% for validation datasampling, which is typical in many studies (due to cost or feasibility issues). Of note, we did notinclude a scenario where both Se and Sp are low, as it may suggest that W is not valid or usefulmeasurement so that ME correction with any method based on these data should beavoided.]

As anticipated, when X is available for all subjects, the results are virtually unbiased(of 0–0.01 bias) and the smallest MSE (of 0.01–0.1) in the log(HR) with accurate CP(0.94–0.96 for almost all scenarios). When we used W for all subjects, the well-known‘attenuation’ or ‘bias toward the null’ phenomenon in the MEliterature with incorrect CP was uniformly observed. When we analyzed the validation sample with Xonly, bias was small (<0.1) but SE and MSE were large due to small sample size. These twonaïve analyses are generally not recommended in practice.

Now we report the performances of different correction methods. MI tended to exhibit thelargest variability among all methods we compared. MI is destined to be unstable when the validationstudy estimator is unstable, e.g., when the size of the validation sample (or the number of events)is too small to result in a reliable imputation model. Overall, RC and Pooled performed comparably,although Pooled was slightly more efficient (i.e., with smaller variance). However, as theorypredicts, when n was large (e.g., n=5000 here), the efficiency gain in Pooled over standardRC was more pronounced (e.g., SSE=0.33 to 0.27). The bias of RC was not systematicallydifferent for common vs. rare events, suggesting that it may be quite robust to the violation of the‘rare event’ assumption. Some portion of bias may have occurred because RC andPooled were originally developed for GLMs (vs. Cox models) with continuous covariates (vs. binarycovariates). Overall, CS tended to provide the smallest bias and the most accurate CP, while RC andPooled tended to provide the smallest MSE. Since CS does not use validation data directly, theresulting estimator was less efficient than RC and Pooled. MC-SIMEX also performed reasonably welland the bias incurred was somewhat comparable to that from RC. It is interesting to note that whenthe true HR was small (i.e., near the null value 1, HR=0.84 in our study), the CP was notextremely low even when W was used for all subjects. When event rate is low (5% here) insmaller total sample size (n=1000), validation sample had only about 5 events so unstableestimation frequently occurred.

Lastly, an important strength of the Cox model is that it can also handle censoringwhich depends on the covariates in the model. Therefore, we repeated the entire simulation under thefollowing setting. We generated time of censoring, C_{new}=I(X=1)*C*0.5+I(X=0)*C*1.5 as a function oftrue covariate X, and C_{new}=I(W=1)*C*0.5+I(W=0)*C*1.5 as a function ofobserved covariate W, where C, X and W were generated as described earlier. For concisepresentation, we reported the results for selected scenarios (i.e., #1, 2, 9, 10, 11 from Table 1) in Table 3.Most interestingly, we observed when censoring time depends on X, RC performed poorly but whencensoring time depends on W, RC performed much better. In contrast, the performance of CS was theopposite, as the theories predicted. Pooled had reduced bias compared to RC in all scenarios. Inthese particular simulations, MI and MC-

SIMEX did not show any noticeable, systematic behaviors.Overall, the performances of the ME correction methods tended to diverge when the censoring was notcompletely random.

4. Application to a Life Course Study with Recalled Childhood SES

In the life course literature, researchers are interested in understanding the potentialeffects of early life experiences on health in later life. While associations between adultsocioeconomic status (SES) and many chronic diseases are well established, the literature on thecontribution of early life SES to the development of chronic diseases in adulthood is lessconclusive. While early life SES is often ascertained via self-report from adults, historicalrecords are regarded as more accurate or objective data sources (Galobardes et al., 2004; Kauhanen et al., 2006).

The Atherosclerosis Risk in Communities (ARIC) study is a prospective study ofcardiovascular disease in a cohort of 15,792 participants from four communities in the US.Recruitment started in 1987–1989 from individuals 45–64 years old. Details aboutthis study have been documented; see http://www.cscc.unc.edu/aric/ and reference(ARIC, 1989). The Life Course SES (LC-SES) study was conductedas an ancillary to ascertain early life SES among over 12,700 ARIC study participants who werecontacted during annual follow-up by telephone in 2001–2002. Details about this study arealso available at http://www.lifecourseepi.info/ and references (Patel etal., 2012; Rose et al., 2008; Rose et al., 2004). Recently, we obtained childhood SES from historicalrecords (e.g., census records) among a sample of participants with the goal of assessing the qualityof recalled early life SES and the impact of the recall error on the association between early lifeSES and adult health outcomes. Specifically, we used the two sources of data (recalled vs.historical) to study the direction and magnitude of the bias in the association of childhood SES andthe two outcomes, mortality and incident coronary heart disease. As a childhood SES measure, we usedfather's occupation and dichotomized non-manual (e.g., professional or managerial) vs.manual occupation groups, which represent 'high SES' vs. 'low SES',respectively. This dichotomization is widely accepted in social, epidemiological and clinicalresearch. In our analysis, 11,264 participants in the original LC-SES study with complete (i.e.,non-missing) recalled SES data and outcomes were included.

Typically, a validation subsample is selected randomly from a full cohort. However, ourvalidation sample was limited to study decedents, as the historical records of interest were onlyaccessible among decedents due to privacy and other administrative reasons. Yet, we do not suspectthat the key assumption of 'non-differential error' was meaningfully violated withthis approach, as there is no data or strong reasons to indicate that decedents would be more orless likely to over- or under-report parental SES than persons who were still alive. Nonetheless,the pooled estimator could be numerically unstable as our validation sample that mostly comprisedevents cannot provide a valid or numerically stable estimate of the HR.

Approximately 16% of the participants had high SES and 6–7% ofthe participants had events. The validation sample showed 54% sensitivity and 86%specificity between recalled data (W) and historical records (X). We found that over-reporting ofSES was more common than under-reporting, which may be interpreted as socially desirable behavior insurveys (Burris et al., 2003). For statistical illustration,we fitted two regression models, a simple regression with the SES variable as a single covariate anda multiple regression adjusting other covariates, where all covariates are time-invariant. Althoughadjusting for intermediate covariates is controversial in life course studies, unadjusted andadjusted models may be justified depending on the goal (Hernandez-Diaz et al., 2006; Oakes and Kaufman,2006).

Table 4 summarizes the regression analyses(i.e., log of HR estimate, SE, and p-value) along with some details about the data and models wefitted.

First, it is noteworthy that de-attenuation by correction methods was not alwaysobserved. For example, MI yielded smaller effect estimates than the naïve estimator in somecases. Moreover, the estimates from MI varied considerably in both magnitude and direction. Ouranalysis highlights that model specifications, which are not always straightforward, especially incomplex real world settings, can be critical for the validity of MI. We observed that RC, CS andMC-SIMEX yielded de-attenuated estimates in all cases; however there was sizable variation in themagnitude. In general, the point estimates from RC and CS were comparable, while those from MC-SIMEXwere closer to the naïve estimates. Overall, CS provided the largest SE, while Pooledprovided the smallest SE. Since the validation data was limited to decedents, our data may not bewell suited for Pooled as mentioned previously. We observed that statistical significance alsovaried across analyses, which may lead to different conclusions. Even within the same method, e.g.,MC-SIMEX, p-value changed somewhat meaningfully depending on the approach used to estimate variance(e.g., asymptotic vs. jackknife method). It is interesting that the effect estimate tended toincrease after ME correction, while the p-value remained similar or increased (Greenland and Gustafson, 2006).

It is important to keep in mind that we intended to deal with *one*statistical problem, ME correction, in this illustrative application. More rigorous investigationsthat address various different issues and aspects of the data are warranted for answering a complexcausal question (Bang, 2010; Greenland, 1980; Greenland and Robins, 1985; Liao et al., 2011; Oakes andKaufman, 2006; Seppa and Hakulinen, 2009).

5. Discussion

In this paper, we compared correction methods for misclassified covariates in the Coxmodel by simulation and data analysis. Our work may be viewed as a natural extension of previouswork in this field (Freedman et al., 2008; Messer and Natarajan, 2008). Exposure ME is highly common as many noted, butfrequently ignored when analyzing epidemiologic data and interpreting the study results (Jurek et al., 2006). In applied research, many do not statisticallyassess or adjust for potential bias in the presence of mismeasured covariates/risk factors.Moreover, if adjustment is attempted, only one method is typically implemented, with the methodoften chosen based on the researchers' familiarity with the method, convention in theirfield or training, and/or the availability of software. However, there are several fundamentallydifferent and computationally feasible methods available. Therefore, we strongly recommend thecorrection of ME should be attempted, whenever justifiable and possible. In our study, we usedpublicly available software or computing programs that required generally minoradaptation/modifications. Although currently available programs are written for different platforms(e.g., SAS, R and Fortran), the absence of universally accepted methods and computational issuesshould not be major barriers in applications. Ideally, a statistical model should not be chosenbased on software availability, simplicity of implementation, or tradition. Also, mechanicalapplication of a method, without proper understanding of important issues and the specific contextcould lead us to erroneous analyses or the same repeated mistake. In general, the choice of the MEshould be guided by: type of variables (e.g., continuous vs. categorical covariates, ME in responseor covariate or both), model (e.g., Cox vs. logistic vs. linear regression), and capabilities of thesoftware, in addition to other fundamental issues such as underlying assumptions and modelsrequired, although some methods seem to be robust to some violations.

As we observed, the attenuation of the regression coefficient for the parameter ofinterest is common when the covariate is misclassified but it is not always the case (Yanez et al., 2002). Not only point estimates but also standarderror estimates should be corrected, which have impacts on confidence interval, hypothesis testing,statistical significance, and power/ sample size estimation. We must emphasize that the quality ofthe validation sample seems to be an essential component. Validation data should provide reliableand precise estimates of sensitivity and specificity for all methods and be large enough for mostmethods. We also found that different ME correction methods need different assumptions and couldlead to meaningfully different results. For example, RC method assumes censoring could depend on W,while CS method assumes censoring could depend on X. In practice, it is generally not easy to figureout the true censoring mechanism. Therefore, it may be reasonable and practical for researchers toimplement more than one correction method whenever they can, preferably with some sensitivityanalyses and careful examination of the assumptions entailed, in order to more fully understand theimpact of systematic error. Also, inconsistent results could be better than one incorrectresult.

Acknowledgments

This research was supported by R01-HL081627 from the National Heart, Lung, and Blood Institute.The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported byNational Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C,HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, andHHSN268201100012C). The authors thank the staff and participants of the ARIC study for theirimportant contributions.

References

Akazawa K, Kinukawa N, Nakamura T. A note on the corrected score function corrected formisclassification. Journal of the Japan Statistical Society. 1998; 28:115–123.

ARIC. The ARIC Investigators. The Atherosclerosis Risk in Community (ARIC) study: design and objectives. 1989; 129:687–702.

Armstrong B. Measurement error in generalized linear models. Communications in Statistics, Series B. 1985; 14:529–544.

Bang H. Medical cost analysis: Application to colorectal cancer data from the SEER Medicaredatabase. Contemporary Clinical Trials. 2005; 26:586–597. [PubMed: 16084777]

Bang, H. Introduction to observational studies. In: Faries, D.; Leon, A.; Haro, J.; Obenchain, R., editors. Analysis of Observational Health-Care Data Using SAS. SAS Press Series; Cary, NC: 2010. p. 3-19.

Burris, J.; Johnson, T.; O'Rourke, D. Validating self-reports of socially desirable behaviors. American Statistical Association Proceedings, American Association for Public OpinionResearch - Section on Survey Research Methods; 2003. p. 32-36.

Carroll, R. Encyclopedia of Biostatistics. Wiley; 2005. Measurement error in epidemiologic studies.

Carroll, R.; Ruppert, D.; Stefanski, L. Measurement Error in Nonlinear Models. Chapman & Hall; London: 1995.

Carroll R, Stefanski L. Approximate quasilikelihood estimation in models with surrogatepredictors. Journal of the American Statistical Association. 1990; 85:652–663.

Cole S, Chu H, Greenland S. Multiple-imputation for measurement-error correction. International Journal of Epidemiology. 2006; 35:1074–1081. [PubMed: 16709616]

Cook J, Stefanski L. A simulation extrapolation method for parametric measurement errormodels. Journal of the American Statistical Association. 1995; 89:1314–1328.

Cox D. Regression models and life-tables (with Discussion). Journal of the Royal Statistical Society, Series B. 1972; 34:187–220.

Dalen I, Buonaccorsi J, Laake P, Hjartåker A, Thoresen M. Regression analysis with categorized regression calibrated exposure: some interestingfindings. Emerging Themes in Epidemiology. 2006; 3:6. [PubMed: 16820052]

Freedman LS, Midthune D, Carroll RJ, Kipnis V. A comparison of regression calibration, moment reconstruction and imputation foradjusting for covariate measurement error. Statistics in Medicine. 2008; 27:5195–5216. [PubMed: 18680172]

Fuller, W. Measurement Error Models. John Wiley & Sons; New York: 1987.

Galobardes B, Lynch J, Smith G. Childhood socioeconomic circumstances and cause-specific mortality in adulthood:systematic review and interpretation. Epidemiologic Reviews. 2004; 26:7–21. [PubMed: 15234944]

Gleser, LJ. Improvements of the naive approach to estimation in nonlinear errors-invariablesregression models. In: Brown, P.; Fuller, W., editors. Statistical Analysis of Measurement Error Models and Applications. American Mathematics Society; Providence: 1990.

Greenland S. The effect of misclassification in the presence of covariates. American Journal of Epidemiology. 1980; 112:564–569. [PubMed: 7424903]

Greenland S, Gustafson P. Accounting for independent nondifferential misclassification does not increasecertainty that an observed association is in the correct direction. American Journal of Epidemiology. 2006; 164:63–68. [PubMed: 16641307]

Greenland S, Robins JM. Confounding and misclassification. American Journal of Epidemiology. 1985; 122:495–506. [PubMed: 4025298]

Hernandez-Diaz S, Schisterman E, Hernan M. The birth weight “paradox” uncovered? American Journal of Epidemiology. 2006; 164:1115–1120. [PubMed: 16931543]

Huang Y, Wang C. Consistent function methods for logistic regression with errors incovariates. Journal of the American Statistical Association. 2001; 95:1209–1219.

Jurek A, Maldonado G, Greenland S, Church T. Exposure-measurement error is frequently ignored when interpreting epidemiologicstudy results. European Journal of Epidemiology. 2006; 21:871–876. [PubMed: 17186399]

Kalbfleisch, J.; Prentice, R. The Statistical Analysis of Failure Time Data. Wiley; New York: 2002.

Kauhanen L, Lakka HM, Lynch J, Kauhanen J. Social disadvantages in childhood and risk of all-cause death and cardiovasculardisease in later life: a comparison of historical and retrospective childhoodinformation. International Journal of Epidemiology. 2006; 35:962–968. [PubMed: 16556645]

Kuchenhoff H, Lederer W, Lesaffre E. Asymptotic variance estimation for the misclassification simex. Computational Statistics and Data Analysis. 2007; 51:6197–6211.

Kuchenhoff H, Mwalili S, Lesaffre E. A general method for dealing with misclassification in regression: Themisclassification SIMEX. Biometrics. 2006; 62:85–96. [PubMed: 16542233]

Lederer W, Kuchenhoff H. A short introduction to the SIMEX and MCSIMEX. R News. 2006; 6:26–31.

Liao X, Zucker DM, Li Y, Spiegelman D. Survival analysis with error-prone time-varying covariates: a risk set calibrationapproach. Biometrics. 2011; 67:50–58. [PubMed: 20486928]

Lindsey J. Fitting parametric counting processes by using log-linear models. Applied Statistics. 1995; 44:201–212.

Little, R.; Rubin, D. Statistical Analysis with Missing Data. John Wiley & Sons; New York: 2002.

Loomis D, Richardson DB, Elliott L. Poisson regression analysis of ungrouped data. Occupational and Environmental Medicine. 2005; 62:325–329. [PubMed: 15837854]

Messer K, Natarajan L. Maximum likelihood, multiple imputation and regression calibration for measurementerror adjustment. Statistics in Medicine. 2008; 27:6332–6350. [PubMed: 18937275]

Nakamura T. Corrected score function of errors-in-variables models: methodology and applicationto generalized linear models. Biometrika. 1990; 77:127–137.

Oakes, J.; Kaufman, J. Methods in Social Epidemiology. Jossey-Bass, A Wiley Imprint; San Francisco, CA: 2006.

Patel MD, Rose KM, Owens CR, Bang H, Kaufman JS. Performance of automated and manual coding systems for occupational data: A casestudy of historical records. American Journal of Industrial Medicine. 2012; 55:228–231. [PubMed: 22420026]

Prentice R. Covariate measurement errors and parameter estimation in a failure time regressionmodel. Biometrika. 1982; 69:331–342.

Qi L, Wang YF, He Y. A comparison of multiple imputation and fully augmented weighted estimators for Coxregression with missing covariates. Statistics in Medicine. 2010; 29:2592–2604. [PubMed: 20806403]

Rose K, Perhac JS, Bang H, Heiss G. Historical records as a source of information for childhood socioeconomic status:results from a pilot study of decedents. Annals of Epidemiology. 2008; 18:357–363. [PubMed: 18395465]

Rose KM, Wood JL, Whitsel EA, Pollitt R, Diez Roux AV, Yoon DK, Knowles S, Heiss G. Linking historical addresses with census tract data from the 1960–80decennial censuses: experiences from the life course SES, social context and cardiovascular diseasestudy. International Journal of Health Geographics. 2004; 17:27. [PubMed: 15548332]

Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervalsfor systematic within-person measurement error (with Discussion). Statistics in Medicine. 1989; 8:1051–1069. [PubMed: 2799131]

Rubin D. Inference and missing data. Biometrika. 1976; 63:581–692.

Seppa K, Hakulinen T. Mean and median survival times of cancer patients should be corrected for informativecensoring. Journal of Clinical Epidemiology. 2009; 62:1095–1102. [PubMed: 19251392]

Slate EH, Bandyopadhyay D. An investigation of the MC-SIMEX method with application to measurement error inperiodontal outcomes. Statistics in Medicine. 2009; 28:3523–3538. [PubMed: 19902495]

Spiegelman D. Regression calibration method for correcting measurement error bias in nutritionepidemiology. American Journal of Clinical Nutrition. 1997; 65:1179S–1186S. [PubMed: 9094918]

Spiegelman D, Carroll RJ, Kipnis V. Efficient regression calibration for logistic regression in main study/internalvalidation study designs with an imperfect reference instrument. Statistics in Medicine. 2001; 20:139–160. [PubMed: 11135353]

Van Buuren S, Boshuizen H, Knook D. Multiple imputation of missing blood pressure covariates in survivalanalysis. Statistics in Medicine. 1999; 18:681–694. [PubMed: 10204197]

White I. Commentary: Dealing with measurement error: multiple imputation or regressioncalibration? International Journal of Epidemiology. 2006; 35:1081–1082. [PubMed: 16847023]

Yanez ND, Kronmal R, Shemanski L, Psaty B. A regression model for longitudinal change in the presence of measurementerror. Annals of Epidemiology. 2002; 12:34–38. [PubMed: 11750238]

Zucker D, Spiegelman D. Corrected score estimation in the proportional hazards model with misclassifieddiscrete covariates. Statistics in Medicine. 2008; 27:1911–1933. [PubMed: 18219700]

Table 1

Simulation study: True parameter value of log (HR)=0.81, under random censoring

#	P(X=1)	Total n	Event rate	Sensitivity/Specificity	Using accurate X	Using observed W	Validation data only	Multiple imputation	Regression calibration	Pooled estimation	Corrected score	MC-SIMEX
1	40%	2000	20%	0.9/0.7 (under-reporting)	0	−33	1	−12	−9	−7	2	−10
					10/10	10/10	34/34	24/25	16/16	15/15	19/19	15/16
					1	12	12	7	3	3	4	3
					96	12	97	93	91	90	96	94
2	40%	2000	5%	0.9/0.7	1	−30	5	−2	−4	−4	0	−6
					20/20	20/21	69/70	56/57	31/31	29/28	33/34	31/32
					4	13	48	31	10	9	11	46
					96	69	97	97	94	94	96	94
3	40%	2000	20%	0.9/0.9	0	−17	1	−10	−7	−6	1	−1
					10/10	10/10	36/33	20/20	12/12	12/11	13/15	18/17
					1	4	13	5	2	2	2	2
					96	59	93	92	91	90	97	96
4	40%	2000	5%	0.9/0.9	1	−15	5	−1	−4	3	1	1
					21/20	21/20	70/70	46/51	25/23	25/22	25/26	24/25
					4	7	49	21	6	6	6	6
					94	85	97	97	94	92	96	98
5	40%	2000	20%	0.7/0.9 (over-reporting)	0	−31	1	−16	−19	−17	2	−5
					10/10	10/10	34/33	22/23	13/13	13/12	18/20	18/15
					1	11	12	7	5	5	3	3
					95	14	94	88	66	68	98	93
6	40%	2000	5%	0.7/0.9	1	−30	3	−1	−19	−17	1	−1
					19/20	20/20	70/70	54/57	24/24	23/23	34/35	31/30
					4	13	49	29	9	8	12	10
					96	68	98	96	88	88	97	93
7	20%	2000	20%	0.7/0.9	0	−34	0	−16	−7	−6	2	−8
					11/11	11/11	38/37	27/27	19/19	18/17	22/25	19/17
					1	22	14	10	4	4	5	4
					95	10	95	91	91	91	98	94

#	P(X=1)	Total n	Event rate	Sensitivity/Specificity	Using accurate X	Using observed W	Validation data only	Multiple imputation	Regression calibration	Pooled estimation	Corrected score	MC-SIMEX
8	20%	2000	5%	0.7/0.9	1	−32	2	−4	−2	−1	−2	13
					21/21	21/22	72/75	59/62	36/36	33/32	39/38	33/34
					4	15	52	35	13	11	15	13
					95	70	96	97	96	95	97	95
9	40%	1000	20%	0.7/0.9	1	−30	1	−13	−19	−17	0	−7
					15/14	15/14	48/49	33/35	19/19	19/17	25/25	22/21
					2	11	23	13	7	7	6	5
					93	45	97	94	80	80	96	95
10	40%	1000	5%	0.7/0.9	0	−31	−7	−8	−20	−19	4	−1
					32/30	31/29	78/100	72/87	38/36	36/34	54/54	46/45
					10	19	61	52	18	17	29	22
					94	81	94	97	91	90	97	93
11	16%	5000	7%	0.55/0.85	0	−52	−4	−12	−10	−6	0	−21
					13/13	13/13	45/43	41/38	33/33	27/26	36/37	24/25
					2	29	20	18	12	8	13	11
					95	1	96	95	93	94	98	92

Entry in each cell represents mean of bias (1st row), sample standard error/mean ofstandard error estimates (2nd row), mean squared error (3rd row), and coverageprobability (last row) for log(HR). All numbers were multiplied by 100. 1000 simulations wereconducted. 10% of total sample size (n) was selected for validation sample. HR denoteshazards ratio.

Table 2

Simulation study: True parameter value of log (HR)= −0.18, under randomcensoring

#	P(X=1)	Total n	Event rate	Sensitivity/Specificity	Using accurate X	Using observed W	Validation data only	Multiple imputation	Regression calibration	Pooled estimation	Corrected score	MC-SIMEX
1	40%	2000	20%	0.9/0.7 (under-reporting)	0 12/12 1 95	8 12/12 2 89	0 42/41 18 95	0 34/33 12 94	3 18/18 3 95	3 18/16 3 93	1 21/22 4 96	3 20/18 4 93
2	40%	2000	5%	0.9/0.7	0 25/25 6 96	8 24/24 6 94	6 77/88 60 90	3 66/74 44 98	4 36/37 13 96	5 34/34 12 95	−2 46/47 21 97	1 38/37 14 96
3	40%	2000	20%	0.9/0.9	0 12/12 1 96	4 12/12 2 94	−2 43/40 19 94	1 24/25 6 95	2 14/14 2 95	2 14/13 2 92	1 15/15 2 96	2 23/21 3 93
4	40%	2000	5%	0.9/0.9	0 26/25 7 95	5 26/25 7 94	6 76/87 58 91	1 56/69 31 99	3 31/29 10 95	4 30/28 9 93	−3 35/33 12 96	−2 46/46 19 96
5	40%	2000	20%	0.7/0.9 (over-reporting)	0 12/12 1 94	6 13/12 2 90	−2 41/39 17 95	1 29/29 84 94	4 15/15 2 94	4 15/14 2 92	1 18/19 3 96	1 16/18 3 94
6	40%	2000	5%	0.7/0.9	−1 25/25 6 96	6 26/26 7 95	6 78/88 61 89	4 65/74 42 99	4 32/32 10 95	4 31/30 10 95	−1 43/42 19 97	5 45/41 21 93
7	20%	2000	20%	0.7/0.9	0 14/14 2 96	8 13/13 2 91	−6 53/49 28 95	−5 42/40 18 95	2 21/22 4 97	3 20/20 4 94	−2 26/27 7 97	0 24/22 6 94

#	P(X=1)	Total n	Event rate	Sensitivity/Specificity	Using accurate X	Using observed W	Validation data only	Multiple imputation	Regression calibration	Pooled estimation	Corrected score	MC-SIMEX
8	20%	2000	5%	0.7/0.9	−2	7	−8	18	1	7	−7	2
					31/30	28/28	70/92	65/82	46/46	42/42	57/63	45/47
					10	8	50	45	21	18	33	20
					95	94	74	97	96	93	97	93
9	40%	1000	20%	0.7/0.9	0	6	−4	−1	4	4	0	1
					17/17	18/18	59/60	44/46	22/22	22/20	28/28	24/26
					3	4	35	19	5	5	8	6
					95	93	97	97	94	93	95	97
10	40%	1000	5%	0.7/0.9	−3	3	28	21	0	3	−3	−3
					41/38	43/40	78/115	74/103	53/49	50/46	67/68	61/62
					17	19	60	59	28	25	45	37
					95	94	67	95	94	90	99	96
11	16%	5000	7%	0.55/0.85	−1	12	−6	−7	2	5	−5	8
					18/18	16/15	59/62	56/58	38/38	33/32	53/63	30/31
					3	4	35	32	14	11	28	10
					95	87	93	96	95	94	96	93

Entry in each cell represents mean of bias (1st row), sample standard error/mean ofstandard error estimates (2nd row), mean squared error (3rd row), and coverageprobability (last row) for log(HR). All numbers were multiplied by 100. 1000 simulations wereconducted. 10% of total sample size (n) was selected for validation sample. HR denoteshazards ratio.

Table 3

Simulation study: True parameter value of log (HR)=0.81, when censoring depends oncovariate

#	P(X=1)	Total n	Event rate	Sensitivity/Specificity	Using accurate X	Using observed W	Validation data only	Multiple imputation	Regression calibration	Pooled estimation	Corrected score	MC-SIMEX
a. when censoring depends on true covariate X												
1	40%	2000	20%	0.9/0.7 (under-reporting)	0	−53	1	−7	−40	−34	1	−10
					14/13	11/11	45/44	37/38	17/16	17/15	36/51	17/18
					2	29	20	14	19	14	13	4
					96	0	96	96	32	39	95	92
2	40%	2000	5%	0.9/0.7	1	−59	18	4	−47	−41	−5	−4
					30/29	21/21	89/95	70/82	32/31	31/30	63/81	39/37
					9	39	82	49	32	26	40	15
					95	20	83	97	66	70	98	95
9	40%	1000	20%	0.7/0.9	1	−43	−1	−9	−33	−31	1	−5
					19/19	18/17	66/67	53/55	23/22	23/21	44/44	22/23
					4	22	44	29	16	15	19	5
					94	42	96	96	65	68	94	92
10	40%	1000	5%	0.7/0.9	−3	−47	45	36	−39	−33	−8	0
					43/42	39/38	83/118	86/108	48/47	46/45	79/90	56/54
					19	37	89	87	38	32	63	31
					96	78	60	94	90	86	98	97
11	16%	5000	7%	0.55/0.85	−1	−71	−1	−5	−56	−42	−9	−24
					20/20	14/13	62/68	60/65	33/33	32/29	54/112	23/23
					4	52	38	36	42	28	30	11
					95	0	91	96	60	70	93	80
b. when censoring depends on true covariateW												
1	40%	2000	20%	0.9/0.7 (under-reporting)	0	−31	2	−3	−7	−6	52	−9
					12/12	13/13	42/41	33/34	21/20	19/18	51/61	21/20
					1	11	18	11	5	4	53	5
					94	37	95	96	92	91	95	91
2	40%	2000	5%	0.9/0.7	3	−28	20	6	−6	−2	42	0
					25/24	27/28	86/85	75/72	44/43	40/38	83/57	43/41

a. when censoring depends on true covariate X

#	P(X=1)	Total n	Event rate	Sensitivity/Specificity	Using accurate X	Using observed W	Validation data only	Multiple imputation	Regression calibration	Pooled estimation	Corrected score	MC-SIMEX
					6	15	78	57	20	16	87	18
					96	81	90	93	95	95	44	96
9	40%	1000	20%	0.7/0.9	1	−29	−1	−8	−18	−15	12	−8
					15/15	21/20	53/51	45/44	25/25	24/22	33/37	25/23
					2	13	28	21	9	8	12	7
					95	68	97	94	88	88	97	91
10	40%	1000	5%	0.7/0.9	−1	−38	10	9	−28	−20	10	−5
					30/30	46/45	81/100	87/94	56/56	49/49	64/70	48/48
					9	36	67	77	39	28	42	23
					94	92	84	96	96	95	97	97
11	16%	5000	7%	0.55/0.85	−0	−53	−3	−6	−11	−4	−19	−22
					12/13	21/21	43/42	42/41	52/52	34/32	79/100	28/26
					1	33	19	18	28	12	66	13
					95	26	96	95	96	94	81	87

Simulation scenarios and numbers (#) are identical in Tables 1 and 3 except for how censoring time was generated. Here, censoring time wasgenerated as a function of covariate: C_{new}=I(W=1)*C*0.5+I(W=0)*C*1.5 in Table 3a andC_{new}=I(X=1)*C*0.5+I(X=0)*C*1.5 in Table 3b.

Entry in each cell represents mean of bias (1st row), sample standard error/mean ofstandard error estimates (2nd row), mean squared error (3rd row), and coverageprobability (last row) for log(HR).

All numbers were multiplied by 100.

1000 simulations were conducted.

10% of total sample size (n) was selected for validation sample.

HR denotes hazards ratio.

Table 4

Results from LC-SES study: Log of HR (SE) and p-value for the association of high SES andevent

Event	P(X=1)	Event rate	Sensitivity/Specificity	Regression	Using recalled data	Multiple imputation* model1/model2	Regression calibration	Pooled Estimation **	Corrected Score	MC-SIMEX asymptotic/jackknife
Death	16%	7%	0.54/0.86	Simple	−0.16 (0.09) p=0.07	0.02/−0.24 (0.25/0.26) p=0.93/0.37	−0.38 (0.21) p=0.07	−0.10 (0.10) p=0.35	−0.45 (0.28) p=0.11	−0.34 (0.18/0.13) p=0.06/0.009
Death	16%	7%	0.54/0.86	Multiple	−0.03 (0.09) p=0.70	−0.10/−0.19 (0.23/0.25) p=0.66/0.46	−0.11 (0.30) p=0.71	−0.16 (0.12) p=0.17	−0.15 (0.39) p=0.70	−0.05 (0.19/0.15) p=0.78/0.73
CHD	16%	6%	0.54/0.86	Simple	−0.24 (0.08) p=0.004	0.06/−0.11 (0.27/0.19) p=0.84/0.54	−0.55 (0.19) p=0.004	−0.29 (0.14) p=0.04	−0.67 (0.30) p=0.03	−0.51 (0.17/0.12) p=0.002/0.00001
CHD	16%	6%	0.54/0.86	Multiple	−0.19 (0.08) p=0.02	−0.11/−0.09 (0.30/0.18) p=0.73/0.61	−0.63 (0.28) p=0.03	−0.24 (0.17) p=0.17	−0.69 (0.35) p=0.05	−0.40 (0.17/0.14) p=0.02/0.004

n=11,264 subjects were in the full sample and 647 subjects were in the internalvalidation sample.

X= father's occupation during childhood obtained from census records (non-manualvs. manual)

W= recalled data when a person was in middle age

Simple regression model includes child SES as single covariate, while multiple regression modeladjusts other covariates (age, race, gender, smoking (cigarette year), diabetes, hypertension andprevalent coronary heart disease). But for MC-SIMEX, we encountered with computational problem so weomitted one covariate, smoking.

* Imputation model 1 (larger model) included W, event indicator, time of event (in year), age,race, gender, smoking, diabetes, hypertension and prevalent coronary heart disease as covariates,while imputation model 2 (smaller model) included W, event indicator and time of event ascovariates. Inclusion of an interaction of W and event does not change the results materially(Results not shown).

HR denotes hazards ratio and SE denotes standard error.

** Some explanations are provided in Section 4 about why pooled estimation may not be well suitedfor this example.