Correcting for measurement error in a two-phase study of Kaposi's sarcoma

Doctoral Qualifying Oral Exam

Josh Slone

04/12/2024

Contents

1 Introduction

The study of Kaposi's sarcoma (KS) in people living with HIV includes a Cox proportional hazards regression for the survival outcome of time from KS diagnosis to death or censoring, a logistic regression for the binary outcome of prevalence of KS at enrollment, and a Poisson regression for the binary outcome of incidence of KS. Errors occur in the outcomes for each analysis: the death indicator δ^* and time from death or censoring Y^* in the time-to-event analysis and Y^* for the prevalence and incidence analyses. Errors also occur in the covariates for each analysis $\boldsymbol{X}^*$: sex, age at diagnosis, year of diagnosis, ART status, and CD4 value. Derived variables, such as age at diagnosis, include multiple underlying variables with errors. Additionally, inclusion criteria, such as the presence and time of KS diagnosis, are misclassified or error-prone. Errors for analysis variables are detailed in Table 1. These factors add complexity to correcting for measurement error in the KS study. Another covariate, program of IeDEA East Africa or CCASAnet is an error free predictor Z^*. Measurement error or misclassification in outcomes and predictors in these regression models may yield invalid inferences. For example, in logistic regression, β_X^* is attenuated relative to β_X^* when Y has nondifferential misclassification, while the impact on the regression estimate may be biased toward or away from the null when misclassification is differential.[1–3] In generalized linear models and Cox proportional hazards model under common conditions of small β_X or very low event rate, estimates for coefficients of error free predictors, for example program β_Z^*, are also biased with the inclusion of error-prone β_X^* in the regression model, leading to an invalid test of the null hypothesis $\beta_Z = 0$.[3–5] Correcting for issues of measurement error and misclassification, such as these, is the focus of the methodological aspect of the KS study. Accordingly, a probabilistic, internal validation subsample was completed of patient records that is approximately optimal for generalized raking in the incidence analysis. Validated records that are part of the validation subsample and unvalidated records are indicated as $V = 1$ or $V = 0$, respectively. The methods used to account for measurement error in the KS study are multiple imputation, inverse probability weighting, and generalized raking.

Table 1: Characteristics of full EHR cohort and unweighted phase 2 samples, and discrepancies

Variable	Phase 1 (N=270,441)	Phase 2 (N=988)	Proportion in error	Discrepancy
KS diagnosis	1.4%	34%	0.012	PPV=0.971, NPV=0.997
Death	6.7%	11.9%	0.012	PPV=0.965, NPV=0.991
Date of death	2014-02-21 (2012-04-26, 2016-04-28)	2013-08-12 (2012-03-11, 2016-02-25)	0.017	-30 (range -654, 34)
Date of ART initiation	2012-08-05 (2009-06-09, 2015-06-07)	2012-08-03 (2010-07-08, 2015-01-04)	0.077	-11 (range -4012, 1668)
Sex, male	37.7%	44.2%	0.002	PPV=0.995, NPV=1
Date of last visit	2018-05-10 (2014-05-29, 2018-12-18)	2018-11-10 (2014-06-23, 2019-01-30)	0.286	84 (range -2304, 3357)
Age at cohort entry	35 (28.3, 42.6)	34.85 (28.6, 42.1)	0.111	0.2 (range -21, 10)
CD4 at cohort entry	272 (122, 467)	258 (114, 456.25)	0.028	8 (range -18, 833)
Date of first visit in cohort	2012-05-08 (2010-03-14, 2015-03-23)	2012-02-04 (2010-04-15, 2014-10-24)	0.118	10 (range -3162, 3069)

Note: [a] East Africa: one patient record removed from the audit for dropping out prior to 2010, 54 patient records removed for not being audited (from MASAKA, TUMBI, AMPATH, MOROGORO, FACES, IDI, MBARARA, RAKAI); CCASANet: 17 patient records removed for not being audited (from PERU).

2 Methods

2.1 Multiple imputation (MI)

With a validation subsample, of completely observed analysis variables $(\boldsymbol{X}, \boldsymbol{X}^*, Z, Y, Y^*, V)$, measurement error can be considered a problem of missing data.[6–8] The unvalidated records $V = 0$ only include observed (X^*, Z, Y^*, V); the validated analysis variables are missing for $V = 0$. Because the validation subsample is a random sample stratified on observed data, the mechanism for missing data is missing at random (MAR) or by design, e.g. $P(V \mid \boldsymbol{X}, \boldsymbol{X}^*, Z, Y, Y^*) = P(V \mid \boldsymbol{X}^*, Z, Y^*)$.[6] Thus, MI, a common approach for addressing missing data, may be used to correct for measurement error. MI has the additional benefit of being able to deal with more complex and efficient MAR sampling,[9] such as in the KS study. Missing or unrecorded values in the unvalidated records, as well as in the validation subsample, can also be accounted for in a natural way by using MI.[10] MI can be used to account for correlated errors in analysis variables and errors in inclusion criteria.[11] MI can also account for differential measurement error, given Y is used for imputing the unknown X.[8]

An MI approach where the validated data $V = 1$ is modeled based on variables from $V = 0$, and imputed values are retrieved as predictions has been studied in several scenarios. Cole, Chu, and Greenland (2006) introduced MI for measurement error correction, where the authors note the MAR validated values by design which they noted immediately facilitates the use of MI.[7] Their simulations in the case of correctly specified imputation and outcome models showed that in the scenario of a time-to-event outcome with a misclassified binary exposure MI is approximately unbiased and yields approximately correct coverage of 95% confidence intervals, where the estimator is more efficient than the validation subsample alone (complete case from a random sample). The authors use $K = 40$ imputed datasets where X^{imp} is imputed based on a model for $X \mid X^*, Y$, with predictions made using a Gaussian distribution to maintain the uncertainty of the imputed values. They note that bias appeared to be increased when $K < 30$. Specifically, they show that with increasing percentage of records validated, the MI estimator has higher power. With increasing measurement error, lower sensitivity and specificity, MI has lower root mean-squared error than the validation subsample and regression calibration, another measurement error correction approach. MI appears to perform less well with a smaller validation subsample and smaller overall sample size. The steps for repeating the imputation process and using Rubin's rules are noted in the appendix. The authors mention that MI can be used for measurement error correction in the context of other types of outcomes, e.g. logistic regression for a binary outcome, and other types of covariates, e.g. continuous with a linear regression. Messer and Natarajan (2008) applied MI to the problem of measurement error with a binary outcome and mismeasured covariate in the setting of logistic regression.[12]

Shepherd, Shaw, and Dodd (2012) outlined an MI approach for addressing measurement error in the scenario of a correlated error-prone predictor and error-prone continuous outcome in an iterative process (including an error free covariate (X^*, Z, Y)).[10] For $V = 1$ records from a validated subsample, a model for $X \mid X^*, Z, Y$ is fit, and the imputed predictor X^{imp} is retrieved from a random draw from the fitted distribution of X for records $V = 0$. For $V = 1$ records, a model for $Y \mid X^{imp}, Z, Y*$ is fit, and the imputed predictor Y^{imp} is retrieved from a random draw from the fitted distribution of Y for records $V = 0$. The imputed variables for $V = 0$ records are combined with the validated records $V = 1$ records to get completed variables to be used in the multiple imputation analysis model. This process is repeated K times to get K datasets and the combined estimates using Rubin's rules. The authors' simulation studies show that in the context of both randomized trials and observational data, as in the KS study, that MI tends to have lower bias, better coverage, and lower mean-squared error (MSE) with increasing proportion of audited patients. MI performed well in each of these areas as compared to the naive estimator using error-prone data and a moment-based measurement error approach. In particular, with as few as 25 validated records of 1,000 simulated records, the MSE for the MI estimator was lower than that of the naive estimator and coverage appeared nominal by 50 audited patients.

Edwards et al. (2013) applied MI to the case of misclassified binary outcomes.[13] The author's simulations show that the MI estimator is approximately unbiased under the scenario of outcome misclassification with several combinations of sensitivity, specificity, and validation subsample sizes. Bias did decrease as the pro-

portion of participants in the validation subgroup increased. Confidence intervals from the MI estimator maintained appropriate coverage and were more efficient than that of the validation subsample alone. Likewise, the MI estimator had smaller mean squared errors than did analysis limited to validation. Only at the lowest level of sensitivity (0.3) did the statistical power of the MI estimator approach the lower power seen in the validation subsample.

Most recently, Giganti et al. (2020) applied MI to address correlated errors in (X^*, Y^*, δ^*), and the inclusion criteria in a time-to-event analysis.[11] The authors explain that since analysis datasets use derived variables, errors in the original variables lead to errors in derived variables, such that errors are dependent across original and derived variables. Errors in either the original variables or derived variables can lead to incorrect implementation of inclusion or exclusion criteria, e.g. an error in a date that places it outside the time frame of the analysis cohort when it is actually during that time frame leads to incorrect exclusion. The authors employ a time discretized approach, e.g. patient-month, to impute the analysis variables and associated eligibility criteria. This granular approach addresses several of the dependencies in the errors simultaneously For example, imputing an event at a particular month based on that month's covariates addresses the date of the event and whether the event occurred allowing for the MI event indicator and time component, as well as eligibility criteria based on the date of the event. The expanded dataset can remain discretized or be summarized to the patient level for an analysis. In their application, the unvalidated data is shown to be biased, whereas their MI estimates are relatively similar to their validation subsample and fully validated cohort data and are more efficient than the estimates from the validation subsample alone. In simulations from the fully validated cohort data, the MSE for the MI estimator is shown to be lower than for that of the validation subsample alone with higher bias but much lower variance across each sample size selected for validation. The authors note the difficulty in meeting the assumption of a correctly specified imputation model in practice. The simulations illustrate that as the magnitude of the association for an omitted predictor increased, the bias increased, MSE increased, and coverage decreased significantly. As expected, misspecifications impacting covariates with smaller associations tended to result in relatively smaller bias. Model overfitting should also be a consideration, as their simpler imputation model appeared to outperform their more complicated imputation model for smaller validation subsample sizes or fewer events. The authors mention that this approach can be used under more complex and efficient probabilistic sampling schemes that meet the MAR assumption. The setting of derived analysis variables with correlated errors is most similar to that of the KS study.

In relation to the KS study, the computational steps of the MI process for missing CD4 values in phase one and phase two data, as well as the steps for measurement error correction in fitting the respective models (e.g. logistic regression for prevalence of KS at enrollment) are detailed in section six. The general approach for estimating $(\beta_{\boldsymbol{X}}, \beta_Z)$ begins with fitting models for error-prone variables using validated records $V = 1$. $V = 0$ records are filled in with a MI approach to account for uncertainty in the predicted values and uncertainty in the estimates of $(\beta_{\boldsymbol{X}}, \beta_Z)$. $X_1 \mid \boldsymbol{X}^* + Z + Y^*$. X_1^{imp} is predicted from $(\hat{\beta}_{\boldsymbol{X}^*},\ \hat{\beta}_Z,\ \hat{\beta}_Y^*)$ where each imputed value is drawn from a multivariate normal distribution with mean $(\hat{\beta}_{\boldsymbol{X}^*},\ \hat{\beta}_Z,\ \hat{\beta}_Y^*)$ and variance $\boldsymbol{\Sigma}_{\hat{\boldsymbol{X}^*}, \hat{Z}, \hat{Y^*}}$. For records with $V = 1$, validated values are used to complete the total records for the imputed values. This process is repeated for each error-prone variable with each successive model including a predictor for previously imputed variables for $V = 1$ and predictions made for $V = 0$ records, e.g. $X_2 \mid \boldsymbol{X}^*_{-X_1^*} + X_1^{imp} + Z + Y^*$. This successive process of imputations would ideally yield Y^{imp} as a last step from $Y \mid \boldsymbol{X}^{imp} + Z + Y^*$ to account for realistic dependencies between imputed values for error-prone variables.[9,11,14] However, in cases with high accuracy of a binary error-prone outcome, a simpler imputation model may be necessary as in the KS study. For each of K multiply imputed datasets, one analysis model is fit and the estimated coefficients and variances are retrieved from each of them. The results are combined using Rubin's rules.[6] For example for the time from KS diagnosis to death or censoring analysis with the Cox proportional hazards model for X_1^{imp} yielding the overall imputed hazard ratio:

$$exp(\bar{\beta}_1^{imp}) \ = \ exp(K^{-1} \sum_{k=1}^{K} \beta_{1,k}^{imp})$$

The variance estimate for X_1^{imp}:

$$\widehat{Var}(\bar{\beta}_1^{imp}) \;=\; K^{-1}\sum_{k=1}^{K}\widehat{Var}(\beta_{1,k}^{imp}) \;+\; (1+K^{-1})(K-1)^{-1}\sum_{k=1}^{K}(\beta_{1,k}^{imp}-\bar{\beta})^2$$

The complexity of dependent derived variables and original error-prone variables in the KS study make MI a natural method to explore in dealing with measurement error in the KS study. Given that the diagnosis of KS Y^* and timing of KS diagnosis are both outcomes and part of the eligibility criteria in our analyses, using a discretized dataset to impute the outcome and correct for eligibility criteria as Giganti et al. (2020) illustrated was a natural analysis step. For example, Y^* is error-prone and its timing and the timing of it with respect to other variables is dependent on correctness of the data, e.g. since KS diagnosis is error-prone the date of KS diagnosis is error-prone and the outcome of time to death or censoring from KS diagnosis in the mortality analysis would be in error or the prevalence of a KS diagnosis could be incorrect. Additionally, error-prone dates, such as date of birth and date of enrollment, occur in the data. Expanded datasets by patient month were created in the MI process to account for dependencies in variables, timing of variables, and the errors in inclusion criteria. This is detailed further in section six.

There are limitations of MI for measurement error correction. Unbiased estimation for MI and nominal confidence interval coverage relies on properly specifying the imputation model[9,11,15]. As noted in Giganti et al. (2020), avoiding model misspecification is difficult in practice. The issue of model misspecification cannot always easily be accounted for by building more complex models, since validation subsample sizes and event rates may not allow for such complexity without overfitting. Uncongeniality between the imputation and analysis models, because of differing covariates or discretized data, can result in biased variance estimators of MI.[11] Using the correct Robins and Wang variance estimator can be difficult to implement.[11,16] When misclassification is infrequent based on the validation subsample, e.g. high positive predictive value and negative predictive value, imputing binary predictors with logistic regression can lead to convergence issues due to separation. The accuracy of the binary covariate allows for a very simple imputation but is something to consider in practice.

2.2 Inverse probability weighting (IPW)

IPW is an unbiased, design-based estimator (as opposed to MI which requires a correctly specified imputation model). IPW is the traditional method to account for designs or sampling schemes for validation subsamples that are more complex than simple random sampling, e.g. probability of selection for validation varies by record. The IPW or Horvitz-Thompson estimator originated with Horvitz and Thompson (1952).[17] Horvitz and Thompson proposed the general form of population total estimator for a survey sample and a standard error estimate. Because estimates for other population quantities are derived from a population total, the Horvitz-Thompson estimator is part of the foundation of design-based inference.[18] With $\pi_i = P(V_i = 1 \,|\, Y_i^*, \boldsymbol{X}_i^*, Z_i) > 0$, the IPW estimator for regression coefficients $\boldsymbol{\beta}$ can be estimated by solving the weighted score function,[19]

$$\sum_{i=1}^{N}\frac{1}{\pi_i}V_i\frac{\partial}{\partial\boldsymbol{\beta}}logP(Y_i|\boldsymbol{X}_i, Z_i;\boldsymbol{\beta}) \;=\; 0$$

In logistic regression, for example, the weighted log-likelihood is maximized with respect to $\boldsymbol{\beta}$ to get the IPW estimator. A weighted Cox model and weighted generalized linear models are used for the KS study in the spirit of survey estimators. Binder (1992) provides a design-based procedure to estimate the Cox regression parameters and their variances.[20] Binder illustrates that the variance estimator of the weighted Cox model is unbiased via derivation and simulations. Binder (1983) derives and proposes a consistent variance estimator for weighted logistic regression where weights come from a survey.[21] Lumley and Scott (2018) explain that the model-based component of the variance estimator of the weighted logistic regression parameter goes to zero as the sampling fraction decreases and the variance estimator becomes the IPW estimating equation of the previous paragraph.[22]

IPW is a natural method to use for measurement error correction in the context of a two-phase design with a validation subsample. Design-based estimators like IPW are not sensitive to the measurement error structure.[23] IPW can be considered from the perspective of statistical sampling where the π_i are known by design and unequal sampling probabilities can be adjusted for as in Horvitz and Thompson (1952).[17] Then, the weighted estimates can be calculated using the previously mentioned process. IPW can also be considered from the perspective of missing data where restricting to the complete cases can induce bias.[24] IPW can correct this bias when the π_i is known, which is the case with validated data where the unvalidated records are missing by design. In either case, the goal is to use the validated subsample to estimate the regression parameters of interest that would have been estimated if we had validated all records.[25,26]

Design-based estimators allow for more complex sampling schemes that can yield more efficient estimates given the unequal probabilities that are a hallmark of IPW. As long as probabilities of selection are positive, over-sampling cases such as with case-cohort and case-control sampling can be used with IPW.[27,28] Any design with a known probability of selection can be used with design-based methods. Optimal designs, or designs that minimize the variance of the estimator of interest, have also been studies for design-based estimators. The Neyman allocation is the optimal sampling strategy for an IPW estimator with mutually exclusive strata and a fixed total sample size.[26,29] Higher numbers of strata also improve efficient of design-based estimators.[30] The standard deviations of the influence functions for the error-prone regression parameters, which describes the impact of removing observations on the regression parameter of interest, are used in the Neyman allocation to determine individual stratum sample sizes in an optimal design.[18,26] Influence functions will be discussed more in the next section on generalized raking as they can be used as auxiliary variable to calibrate the weights for an IPW estimator.

IPW estimators can be shown to be unbiased given that the sampling weights cancel out the probability of selection in an expectation. This robustness, along with minimal assumptions (MAR sampling without missingness in the validation) make IPW appealing as a first step in correcting for measurement error with a two-phase design. However, despite relatively more efficient sampling schemes mentioned above, IPW estimators can be inefficient given that they ignore most of the data in the cohort.[31,32] Robins et al. (1994) show that the Horvitz-Thompson estimator is inefficient in their class of semiparametric incomplete-data estimators and that estimating the π_i will yield a more precise estimator of the complete data estimating function. The inefficiency of some IPW estimators led to an interest in using generalized raking for two-phase designs for measurement error correction, which can then incorporate unvalidated data to improve the precision of estimates for regression parameters by calibrating the π_i.[33] Generalized raking is efficient in the Robins et al. (1994) class of semiparametric incomplete-data estimators or augmented inverse probability weighted (AIPW) estimators.[30] While the optimal set of auxiliary variables are unobserved, Breslow et al. (2009 in AJE) illustrated generalized raking in practice for a survival outcome using simulations from the National Wilms Tumor Study to show that the IPW, estimation, and generalized raking are all unbiased but the latter methods are more efficient given auxiliary information.[34]

2.3 Generalized raking (GR)

Generalized raking is a design-based estimator that augments the IPW estimator with auxiliary information to improve efficiency. GR makes the smallest adjustments to the π_i from IPW so estimated totals match the totals from auxiliary variables. The auxiliary variables should be known and provide additional information about the relationship being estimated. In this scenario, the calibrating the π_i improves on the precision of the IPW estimator.[18] The goal is to obtain the calibrated weights g_i/π_i and proceed with estimation as in IPW. Minimizing the distance $\sum_{i=1}^{N} V_i d(1/\pi_i, g_i/\pi_i)$ subject to the constraint $\sum_{i=1}^{N} \frac{g_i}{\pi_i} V_i W_i = \sum_{i=1}^{N} W_i$ yields the g_i, where $d(\cdot)$ is a distance function and W is an auxiliary variable.[26] The minimization can be solved by Lagrange multipliers with distance functions outlined in Deville and Sarndal (1992) and Lumley (2010).[35] A reasonable distance function to ensure $g_i/\pi_i > 0$ is $d(a,b) = alog(a/b) + (b-a)$; Deville and Sarndal (1992) show that all GR estimators are asymptotically equivalent.[36] The goal of GR for the KS study is to estimate regression parameters $\boldsymbol{\beta}$. As Breslow et al. (2009) detail, the variance of parameter estimates is the sum of the typical model-based variance from estimation and the design-based variance from IPW estimation of the sum of the influence functions (IFs).[32] This second component of the variance can be

reduced by adjusting the weights via calibration with estimated cohort totals of the IFs. These totals of the estimated influence function serve as the auxiliary variables in GR estimation. Finally, estimation for $\boldsymbol{\beta}$ for GR proceeds by solving the weighted score function,

$$\sum_{i=1}^{N} \frac{g_i}{\pi_i} V_i \frac{\partial}{\partial \boldsymbol{\beta}} logP(Y_i|\boldsymbol{X}_i, Z_i; \boldsymbol{\beta}) \ = \ 0$$

Lumley and Scott (2017) details the mathematical definition of the influence function for a generalized linear model.[22] Oh et al. (2019) details the definition of the IFs for Cox models. Breslow et al. (2009) showed that since the sum of the IFs from the error free cohort estimate the regression slope parameter, IFs for the true population totals are optimal auxiliary variable in GR estimation. Specifically, for IF $\boldsymbol{h}_i(\beta)$, Breslow et al. (2009) showed that $E[\boldsymbol{h}_i(\beta)|Y, X, Z, V]$ is optimal for GR. Of course, these true population totals are not known in the case of error-prone phase one cohort data. Instead Breslow et al. (2009), Lumley (2010), and Oh et al. (2019) have used the error-prone full cohort IFs. These IFs can be highly correlated with the true influence functions under lower levels of measurement error, as in the KS study. Oh et al. (2020) plot the true IFs against mismeasured IFs to show potential departures from linearity for survival data.[36] The efficiency of the GR estimator depends on the correlation between the auxiliary raking variables and the parameter of estimation.[30] With nonlinear relationships or lower correlation, GR can be inefficient in practice.[36] That one can maintain the robustness of the IPW estimator when combined with a precise but potentially inconsistent model-based estimator has general appeal, given IPW's widespread use.[30]

The `dfbetas()` from `resid()` using `R` can be used with negligible error to retrieve the influence functions for a Cox model fit.[36] Similar approximations to the influence functions for generalized linear models can be retrieved from model fits using `R`. For logistic regression model `fit` from Tong Chen:

```
MM <- model.matrix(fit)  Ihat <- (t(MM) %*% (MM * fit$fitted.values * (1 - fit$fitted.values)))
/ nrow(MM)  IF <- (MM * resid(fit, type = "response")) %*% solve(Ihat)
```

For Poisson regression model `fit` from Gustavo Amorim:

```
MM <- model.matrix(fit)  Score.P   <- fit$y*MM - MM*exp(fit$linear.predictors)  InfoMat.P
<- t(MM*sqrt(exp(fit$linear.predictors))) %*% (MM*sqrt(exp(fit$linear.predictors)))/nrow(MM)
IF <- (Score.P) %*% solve(InfoMat.P)
```

IFs from MI estimators from the target regression can be used as auxiliary variables. Oh et al. (2020) mentions a conditional MI approach that iteratively imputes error-prone variables. Though Oh et al. (2020) does not find an appreciable improvement in performance for a survival outcome with misclassified outcome and time-to-event. This approach is necessary in the KS study, given the error-prone inclusion criteria that leads to the phase two analysis cohort not necessarily being a subset of the phase one analysis cohort. For example, a record $V = 1$ may have been a false negative for KS. Using MI IFs with GR is going to be a topic of further study. The process for estimating GR regression coefficients with MI proceeds as follows. First, iteratively impute phase two variables conditional on phase one or imputed version of these variables, e.g. X^{imp} from the MI section. Fit the target model to the imputed variables, including for validated records. Ensure that the validation analysis cohort is a subset of the fully imputed analysis cohort dataset. Retrieve the influence functions with the previously mentioned approaches. Calibrate the weights using `calibrate()` from the `survival` package with option `calfun="raking"`. Estimate $\boldsymbol{\beta}$ using the using the design object with the calibrated weights via `svycoxph()` or `svyglm()`. Combine the estimates and variances as described in the MI section.[32] The variances from GR with MI IFs may be worthy of future study given both the issue of differences in the analysis cohorts and the uncongeniality of the imputation and analysis models.

Breslow et al. (2009) noted that while Robins et al. (1994) showed that increasing the number of auxiliary variables used to estimate g_i/π_i will never increase the asymptotic variance of the estimator, their simulations showed that convergence issues with many raking variables in the calibration algorithm can lead to larger standard errors in finite samples. Estimated the sampling weights using the MI data with logistic regression may be an approach to consider in situations with many predictors of interest and smaller validation subsample sizes. Oh et al. (2019) showed via simulations the finite sample properties of the GR

estimator for a mismeasured survival outcome.[23] The GR estimator had small bias and the smallest MSE of the estimators in most of the scenarios studied. Scenarios with moderate to larger hazard ratios and more common events yielded the best performance from the GR estimator. As noted, Oh et al. (2019) simulations did not show a difference between using the error-prone IFs and IFs using a direct imputation approach. Oh et al. (2020) studied survival outcomes in the presence of correlated measurement error of (X^*, Y^*, δ^*) using GR estimators.[36] They show that the GR estimator with error-prone IFs is inefficient when δ_i is misclassified. Their general approach is to construct auxiliary variables that are more highly correlated with true population IFs. They proceed by combining the GR estimator with fully conditional MI to iteratively impute the error-prone variables and retrieve the IFs. This process resulted in higher relative efficiency compared to the IPW estimator and the GR estimator with error-prone IFs in nearly all simulated scenarios. With much higher positive predictive values and negative predictive values, the MI IFs yield larger efficiency gains. Han (2016) showed that for design-based estimators consistency does not rely on a correctly specified imputation model.[37] The efficiency gains of the GR estimator will be less appreciable than those of a correctly specified MI modeling procedure.[36] Considering the complexity of large cohorts from observational databases, a correctly specified MI modeling procedure is difficult in practice. Han et al. (2020) demonstrated that minimal misspecification of the MI model results in bias and worse MSE than the GR estimator.[38] With minimal misspecification of the MI model, Han et al. (2020) show that the GR estimator has a better robustness-efficiency trade-off than MI alone.[38]

The GR estimator is included in the class of regular asymptotically linear estimators consistent for the design-based parameter of interest or AIPW estimators.[31] The class of GR estimators include the most efficient AIPW estimators, so that GR estimators are asymptotically efficient among design-based estimators.[30] The GR estimators from Oh et al. (2020) that utilize MI allow for approximations of the optimal AIPW estimator in measurement error scenarios that involve dependencies and derived variables that often occurs in observational data.[36] For measurement error setting where the MI IFs are correlated with the true population IF, Oh et al. (2020) show that their estimators are robust to imputation and outcome model misspecification. Under correct specification, these GR estimators yield the most efficient design-based estimator. For instances of lower correlation of these IFs, the GR estimator may be no more efficient than the IPW estimator. The improvements in design-based efficiency from GR are appealing for measurement error correction in two-phase designs. Nonetheless, due to the potential of overfitting and finite sample issues in calibration convergence for larger number of raking variables, as discussed in[32], the size of the validation subsample or number of events should be considered in practice.

2.4 MI approach

Initially impute whether patients with error-prone KS from 2010-2019 remain in the study. No individuals from CASSAnet had a diagnosis changed to outside the study period. Thus, we stratified by program and imputed exclusion for these patients using the probability of this occurrence in the validation subsample for Bernoulli draws.

1. Imputed phase one birth dates

- Logistic regression prediction model from phase two data with an outcome indicating whether birth date is in error
- Linear regression prediction model from phase two data with an outcome of number of days in error from validated birth date for records having an error in birth date
- Calculated fully imputed birth dates (with imputed values for $V = 1$) and partially imputed birth dates (with validated values for $V = 1$)

2. Imputed phase one earliest date in the data, which is based on enrollment date or earlier CD4 visit date

- Logistic regression prediction model from phase two data with an outcome indicating whether the earliest date is in error

- Linear regression prediction model from phase two data with an outcome of number of days in error from validated earliest date for records having an error in earliest date
- Calculated imputed earliest date (from added or subtracted days for those records predicted in error)

3. Imputed phase one latest date in the data, which is based on the most recent of last alive date, last visit date, and death date

- Logistic regression prediction model from phase two data with an outcome indicating whether the latest date is in error
- Linear regression prediction model from phase two data with an outcome of number of days in error from validated latest date for records having an error in their latest date
- Calculated imputed latest date. $<<$ 1% of the records were manually changed to include a latest data equal to the earliest date in cases where it was predicted to be prior to the earliest date

3. Imputed phase one enrollment for those that have a validated enrollment date after a validated earliest date in the data (limited to instances of an earlier CD4 visit that occurs in $\sim$ 5% of PHW records)

- Logistic regression prediction model from phase two data with an outcome indicating whether the enrollment date is in error
- Linear regression prediction model from phase two data with an outcome of number of days in error from validated enrollment date for records having an error in their enrollment date
- Calculated imputed enrollment date. $<<$ 1% of the records were manually changed to include a latest date equal to the enrollment date in cases where enrollment date was predicted to be after the latest date

5. Dataset is expanded by patient-month from the earliest date in the data to the latest date

6. CD4 is carried forward 12 months and carried backwards within six months of the earliest date. Missing values of CD4 in both phase one and phase two that still exists after this process are indicated for imputation

7. Imputed sex stratified by program East Africa (EA) vs. CASSAnet

- Bernoulli draws with positive predictive value (PPV = 355/357, so some males from EA are predicted to be female as in phase two and no males change from EA) and negative predictive value (NPV=1 for both cohorts, so no females change to male) for validated male indicator from phase two
- There are fully imputed and partially imputed variables for sex

8. Imputed KS indicator stratified by program EA vs. CASSAnet

- Bernoulli draws with EA PPV and NPV (11048/11230 and 34478/34586, respectively) and CASSAnet PPV and NPV (1590/1710 and 4909/4918, respectively) for validated KS indicator from phase two data
- There are fully imputed and partially imputed variables for the KS indicator
- All rows (patient-month) after first predicted KS have KS indicator of one

9. Imputed phase one CD4 values based on error-prone phase one variables to address phase one CD4 missingness

- Linear regression prediction model from phase one square root CD4 value continuous variable conditional on phase one predictors

- Use imputed values to fill in missing values (indicator from 6. above) and carry forward imputed CD4 values for 12 months
- << 1% of the records with imputed phase one CD4 less than zero were assigned a one

10. Imputed phase two CD4 values based on validated variables to address phase two CD4 missingness

- Linear regression prediction model from phase two square root CD4 value continuous variable conditional on phase two predictors
- Use imputed values to fill in missing values (indicator from 6 above) and carry forward imputed CD4 values for 12 months

11. Imputed phase one CD4 values based on validated variables to address error-prone phase one CD4

- Linear regression prediction model from phase two square root CD4 value continuous variable conditional on imputed phase one CD4 (from 9 above) and imputed phase one variables and other error-prone phase one variables
- Use imputed values for error-corrected phase one CD4 values for all records for fully imputed and for where $V = 0$ for partially imputed

12. Imputed phase one ART indicator

- Logistic regression prediction model from phase two data with an outcome indicating whether the patient was on ART conditional on phase one ART indicator and imputed phase one variables and other error-prone phase one variables
- There are fully imputed and partially imputed versions of the ART indicator
- All rows after first predicted ART indicator of one have a value of one

13. Imputed phase one death indicator

- Logistic regression prediction model from phase two death indicator variable as the outcome conditional on phase one death indicator and imputed phase one variables
- Predictions are made for the last month of each patient's expanded data
- There are fully imputed and partially imputed versions of the ART indicator

2.5 Analyses

2.5.1 Time from KS diagnosis to death or censoring

2.5.1.1 Creation of time from KS diagnosis to death or censoring cohort

The patient-month expanded and imputed dataset is grouped to the patient level. The death date or censoring date are determined and added to every row. Only the rows (month) where a patient was imputed as being diagnosed with KS are selected. Analysis variables are created.

A similar process is used to get the fully imputed cohort used for GR and the EHR cohort.

2.5.1.2 Analysis approaches for time from KS diagnosis to death or censoring

MI to correct for measurement error is fit using a Cox proportional hazards regression with an outcome of partially imputed (validated values from $V = 1$ records are not imputed) time from KS diagnosis to death or censoring and an indicator for death or censoring. Predictors include partially imputed sex, age at diagnosis, year of diagnosis, ART status at diagnosis, square root CD4 value, and error free program using `coxph()`.

EHR or the naive analysis, based on the phase one data, includes the same model with error-prone versions of variables used in the multiple imputation model directly above, including square root CD4 from the phase one imputation using `coxph()`.

IPW is fit using the data limited to phase two, attaching the sampling weights for the `twophase()` function in the `survey` package in `R`, and fitting a weighed version of the model described above but using the validated values only in the `svycoxph()` function with the design from `twophase()`.

GR analysis begins with the Cox model fit of the fully imputed (validated values from $V = 1$ records are imputed) phase one data that only includes patients who were also diagnosed with KS in phase two. 1) This limitation may remove patients with validated KS diagnoses, because they may be imputed to not have a KS diagnosis in the phase one fully imputed data. 2) The phase two data used in the GR weighted Cox model is limited to those with a validated KS diagnosis. The fully imputed phase one data may include patients who have a KS diagnosis but were validated to not have a KS diagnosis. Thus, records may be removed from the phase two data for not being imputed to have a KS diagnosis (used in the weighted GR analysis), and records may be removed from the fully imputed phase one data (used to fit the Cox model used to retrieve the influence functions) because they did not have a validated KS diagnosis. The fully imputed phase one data used to fit the Cox model to retrieve the influence functions is created by limiting the dataset to rows at KS diagnosis, and the associated model is fit. The influence functions from the model for the fully imputed variables are retrieved with `resid() type="dfbeta"`, and the influence functions for all predictors are attached to the imputed dataset. The twophase() function from the `survey` package is used to set up the two phase design, denote the weights, and subset to $V = 1$. The `calibrate()` function is used to calibrate the weights via generalized raking. The `svycoxph()` functions fits the weighted Cox model with the design and calibrate weights from `twophase()` and `calibrate()`. The validated variables are used in this model fit.

This process is repeated for each imputed dataset and the coefficient estimates and standard error estimates are combined using `MIcombine()` (Rubin's rules).

2.5.2 Prevalence of KS at enrollment or cohort entry

2.5.2.1 Creation of prevalence cohort

The patient-month expanded and imputed dataset is grouped to the patient level. Those with imputed enrollment date prior to 2010 are removed. The prevalence indicator is determined by assigning those without a KS diagnosis to zero and assigning those who were imputed to have been diagnosed with KS within 60 days of imputed enrollment date a one. Analysis variables are created.

A similar process is used to get the fully imputed cohort used for GR and the EHR cohort.

2.5.2.2 Analysis approaches for prevalence of KS

MI to correct for measurement error is done using a logistic regression with an outcome of the indicator for prevalence. Predictors include partially imputed sex, age at diagnosis, year of diagnosis, ART status at diagnosis, square root CD4 value, and error free program using `glm()`.

EHR or the naive analysis, based on the phase one data, includes the same model and error-prone versions of variables used in the multiple imputation model directly above, including square root CD4 from the phase one imputation using `glm()`.

IPW is done using the data limited to phase two, attaching the sampling weights for the `twophase()` function in the `survey` package, and fitting the weighting model described above but using the validated values only in the `svyglm()` function with the design from `twophase()`.

GR analysis begins with the logistic regression model fit of the fully imputed (validated values from $V = 1$ records are imputed) phase one data. The fully imputed phase one data used to fit the logistic model to retrieve the influence functions is created, and the associated model is fit. The influence functions from the model for the fully imputed variables are retrieved with Tong's function, and the influence functions for all predictors (excluding the intercept) are attached to the imputed dataset. The `twophase()` function from the `survey` package is used to set up the two phase design, denote the weights, and subset to $V = 1$. The `calibrate()` function is used to calibrate the weights via generalized raking. The `svyglm()` functions fits the weighted logistic regression model with the design and calibrate weights from `twophase()` and `calibrate()`. The validated variables are used in this model fit.

This process is repeated for each imputed dataset and the coefficient estimates and standard error estimates are combined using `MIcombine()` (Rubin's rules).

2.5.3 Incidence of KS over the study period

2.5.3.1 Creation of incidence cohort

The patient-month expanded and imputed dataset is placed in a counting process format, primarily following the changes in CD4 values by patients over time. The event indicator for incidence is determined by assigning those without a KS diagnosis a zero. Those who were imputed to have been diagnosed with KS at least 60 days after the imputed enrollment date a one. Those who were imputed to have been diagnosed with KS within 60 days of imputed enrollment date are removed. All rows for prevalent cases are removed; all rows for those with no imputed KS diagnosis are maintained, and only those rows until KS diagnosis are maintained for those who had an imputed diagnosis. Analysis variables are created.

A similar process is used to get the fully imputed cohort used for GR and the EHR cohort.

2.5.3.2 Analysis approaches for incidence of KS

MI to correct for measurement error is done using a Poisson regression with an outcome of the event of KS diagnosis indicator and an offset of 1,000 person-years. Predictors include partially imputed sex, age at diagnosis, year of diagnosis, ART status at diagnosis, square root CD4 value, and error free program using `glm()`.

EHR or the naive analysis, based on the phase one data, includes the same model and error-prone versions of variables used in the multiple imputation model directly above, including square root CD4 from the phase one imputation using `glm()`.

IPW is done using the data limited to phase two, attaching the sampling weights for the `twophase()` function in the `survey` package, and fitting the weighting model described above but using the validated values only in the `svyglm()` function with the design from `twophase()`.

GR analysis begins with the Poisson regression model fit of the fully imputed (validated values from $V = 1$ records are imputed) phase one data. The influence functions from the model for the fully imputed variables are retrieved with Gustavo's function that multiplies the score matrix by the inverse of the information matrix. The rows for each influence function for all predictors receive an ID of its associated patient-year combination. The data is grouped by patient year, and each of the influence functions is summed for each

combination. The sum totals of the number of rows in the fully imputed dataset and the sum totals of the influence functions are calculated to be used as the auxiliary variables for GR. The validated data is limited to patients with patient-year combinations from the fully imputed data so that each row has an associated influence function. This is necessary due to the requirement that the phase two data include the auxiliary variables for generalized raking with the `survey` package. The patient-year combinations are used since the group of patients may differ between the fully imputed phase one data and the phase two data, due to changes in timing of KS diagnosis prior to enrollment leading to inclusion of additional prevalent cases. These prevalent cases are removed from the data as part of the inclusion criteria, so at minimum the data should be grouped by patient. Additionally, though, the counting process data also yields differing number of rows based on changes in CD4 values, which occur more frequently in the fully imputed data where the CD4 value is imputed by row whereas the missing validated CD4 values are carried forward for 12 months. Because CD4 values are time dependent, using patient-year combinations preserves some of the granularity of changes in CD4. The summed influence functions for patient-year combinations are attached to the validated data. The `svydesign()` function from the `survey` package is used to set up the two phase design, denoting the weights for the validated data. The `calibrate()` function is used to calibrate the weights via generalized raking using the population totals. The `svyglm()` functions fits the weighted Poisson model with the design and calibrate weights from `svydesign()` and `calibrate()`. The validated variables are used in this model fit.

This process is repeated for each imputed dataset and the coefficient estimates and standard error estimates are combined using `MIcombine()` (Rubin's rules).

2.6 Proposed work (next steps)

The issue of error-prone eligibility arose during the analysis for the KS study. There were false positives and false negatives for KS diagnosis. There was an instance of a patient's date of KS diagnosis being correct such that they were no longer eligible for the study. The date of diagnosis was corrected to pre-2010. The cohort is limited to those without KS from 2010-2019 and to those who were diagnosed with KS from 2010-2019. The issue with eligibility criteria impacts both the entire cohort used for the MI process, to address missing CD4 values and measurement error correction. MI can account for errors in eligibility criteria as shown by Giganti et al. (2020). These methods are employed in the KS study for the MI measurement error correction approach. The eligibility criteria also differ by analysis. The time from KS diagnosis to death or censoring analysis requires a KS diagnosis from 2010-2019. The prevalence of KS analysis at enrollment requires enrollment during 2010-2019. The incidence of KS analysis requires the removal of prevalent cases. Errors in eligibility criteria that impact the entire cohort can be accounted for with MI and used in GR. The individual analyses, however, yield instances where phase two is no longer a subset of phase one in the analysis cohorts, which differs from the typical two-phase design. For the mortality analysis, the patients without validated KS diagnosis are simply removed. For the accompanying GR where fully imputed influence functions are used, patients differ between phase one and phase two because of the imputed KS indicator for all patients (including phase two). For the incidence analysis, GR where fully imputed influence functions are used, patients differ between phase one and phase two because of the imputed KS indicator for all patients (including phase two). Some patients are imputed to have KS before or within 60 days of their enrollment data and are removed. Other patients who were prevalent cases based on the validated data are no longer prevalent cases because of the imputed KS indicator. The validation cohort for these analyses is no longer a subset of the fully imputed phase one data. This issue can arise when using error-prone influence functions as well, because some control patients may become cases and so would be excluded from the phase one data but be included in the validation analysis cohort. The issue of error-prone inclusion criteria while using the GR estimator has room for further exploration. We plan to explore this beginning with a simulation study to understand the impact of differing levels of error-prone eligibility criteria on the GR estimator. We plan to also explore whether there are difference in performance between using the average of the influence functions from each MI dataset for one GR estimation and averaging the GR estimates from MI datasets. This will be a step toward determining the impact of error-prone eligibility criteria on the bias and variance of the GR estimator.

References

1. Magder LS, Hughes JP. Logistic Regression When the Outcome Is Measured with Uncertainty. *American Journal of Epidemiology*. Published online 1997. doi:10.1093/oxfordjournals.aje.a009251
2. Buonaccorsi JP. *Measurement Error: Models, Methods, and Applications*. Chapman; Hall/CRC; 2010. doi:10.1201/9781420066586
3. Keogh RH, Shaw PA, Gustafson P, et al. STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1-Basic theory and simple methods of adjustment. *Statistics in Medicine*. Published online 2020. doi:10.1002/sim.8532
4. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*. Published online 1982. doi:10.1093/biomet/69.2.331
5. Xie SX, Wang CY, Prentice RL. A risk set calibration method for failure time regression by using a covariate reliability sample. *Journal of The Royal Statistical Society Series B-statistical Methodology*. Published online 2001. doi:10.1111/1467-9868.00317
6. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. 1st ed. Wiley; 2002. doi:10.1002/9781119013563
7. Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. *International Journal of Epidemiology*. Published online 2006. doi:10.1093/ije/dyl097
8. Shaw PA, Gustafson P, Carroll RJ, et al. STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 2—More complex methods of adjustment and advanced topics. *Statistics in Medicine*. Published online 2020. doi:10.1002/sim.8531
9. Shepherd BE, Shaw PA. Errors in multiple variables in human immunodeficiency virus (HIV) cohort and electronic health record data: Statistical challenges and opportunities. *Statistical Communications in Infectious Diseases*. Published online 2020. doi:10.1515/scid-2019-0015
10. Shepherd BE, Shaw PA, Dodd LE. Using audit information to adjust parameter estimates for data errors in clinical trials. *Clinical Trials*. Published online 2012. doi:10.1177/1740774512450100
11. Giganti MJ, Shaw PA, Chen G, et al. ACCOUNTING FOR DEPENDENT ERRORS IN PREDICTORS AND TIME-TO-EVENT OUTCOMES USING ELECTRONIC HEALTH RECORDS, VALIDATION SAMPLES, AND MULTIPLE IMPUTATION. *The Annals of Applied Statistics*. Published online 2020. doi:10.1214/20-aoas1343
12. Messer K, Natarajan L. Maximum likelihood, multiple imputation and regression calibration for measurement error adjustment. *Statistics in Medicine*. Published online 2008. doi:10.1002/sim.3458
13. Edwards JK, Cole SR, Troester MA, Richardson DB. Accounting for Misclassified Outcomes in Binary Regression Models Using Multiple Imputation With Internal Validation Data. *American Journal of Epidemiology*. Published online 2013. doi:10.1093/aje/kws340
14. Buuren S van. *Flexible Imputation of Missing Data*. Chapman; Hall/CRC; 2012. doi:10.1201/b11826
15. McIsaac MA, Cook RJ. Statistical methods for incomplete data: Some results on model misspecification. *Statistical Methods in Medical Research*. Published online 2017. doi:10.1177/0962280214544251
16. Robins JM, Wang N. Inference for imputation estimators. *Biometrika*. Published online 2000. doi:10.1093/biomet/87.1.113
17. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*. Published online 1952. doi:10.1080/01621459.1952.10483446
18. Lumley T. *Complex Surveys: A Guide to Analysis Using r*. 1st edition. Wiley; 2010.
19. Chen T, Lumley T. Optimal sampling for design-based estimators of regression models. *Statistics in Medicine*. Published online 2022. doi:10.1002/sim.9300

20. Binder DA. Fitting Cox's proportional hazards models from survey data. *Biometrika*. Published online 1992. doi:10.1093/biomet/79.1.139

21. Binder DA. On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*. Published online 1983. doi:10.2307/1402588

22. Lumley T, Scott A. Fitting Regression Models to Survey Data. *Statistical Science*. Published online 2017. doi:10.1214/16-sts605

23. Oh EJ, Shepherd BE, Lumley T, Shaw PA. Raking and Regression Calibration: Methods to Address Bias from Correlated Covariate and Time-to-Event Error. *arXiv: Methodology*. Published online 2019. doi:10.1002/sim.8793

24. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*. Published online 2013. doi:10.1177/0962280210395740

25. Lawless JF, Kalbfleisch JD, Wild CJ. Semiparametric methods for response-selective and missing data problems in regression. *Journal of The Royal Statistical Society Series B-statistical Methodology*. Published online 1999. doi:10.1111/1467-9868.00185

26. Amorim G, Tao R, Lotspeich SC, Shaw PA, Lumley T, Shepherd BE. Two-phase sampling designs for data validation in settings with covariate measurement error and continuous outcome. *Journal of The Royal Statistical Society Series A-statistics in Society*. Published online 2021. doi:10.1111/rssa.12689

27. White JE. A TWO STAGE DESIGN FOR THE STUDY OF THE RELATIONSHIP BETWEEN A RARE EXPOSURE AND A RARE DISEASE. *American Journal of Epidemiology*. Published online 1982. doi:10.1093/oxfordjournals.aje.a113266

28. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. Published online 1986. doi:10.1093/biomet/73.1.1

29. Neyman J. On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society*. Published online 1934. doi:10.1007/978-1-4612-4380-9_12

30. Lumley T, Shaw PA, Dai JY. Connections between Survey Calibration Estimators and Semiparametric Models for Incomplete Data. *International Statistical Review*. Published online 2011. doi:10.1111/j.1751-5823.2011.00138.x

31. Robins JM, Rotnitzky A, Zhao LP. Estimation of Regression Coefficients When Some Regressors are not Always Observed. *Journal of the American Statistical Association*. Published online 1994. doi:10.1080/01621459.1994.10476818

32. Breslow NE, Lumley T, Ballantyne CM, Chambless LE, Kulich M. Improved Horvitz-Thompson Estimation of Model Parameters from Two-phase Stratified Samples: Applications in Epidemiology. *Statistics in Biosciences*. Published online 2009. doi:10.1007/s12561-009-9001-6

33. Deville JC, Särndal CE, Sautory O. Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*. Published online 1993. doi:10.1080/01621459.1993.10476369

34. Breslow NE, Lumley T, Ballantyne CM, Chambless LE, Kulich M. Using the Whole Cohort in the Analysis of Case-Cohort Data. *American Journal of Epidemiology*. Published online 2009. doi:10.1093/aje/kwp055

35. Deville JC, Särndal CE. Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*. Published online 1992. doi:10.1080/01621459.1992.10475217

36. Oh E, Shepherd BE, Lumley T, Shaw PA. Improved Generalized Raking Estimators to Address Dependent Covariate and Failure-Time Outcome Error. *Biometrical journal Biometrische Zeitschrift*. Published online 2020. doi:10.1002/bimj.202000187

37. Han P. Combining inverse probability weighting and multiple imputation to improve robustness of estimation. *Scandinavian Journal of Statistics*. 2016;43(1):246-260. doi:10.1111/sjos.12177

38. Han K, Shaw PA, Lumley T. Combining multiple imputation with raking of weights: An efficient and robust approach in the setting of nearly-true models. *Statistics in Medicine*. Published online 2020. doi:10.1002/sim.9210