Title

Estimating multiple time-fixed treatment effects using a semi-Bayes semiparametric marginal structural Cox proportional hazards regression model

Journal

Biometrical Journal, 60(1)

ISSN

0323-3847

Authors

Cole, Stephen R
Edwards, Jessie K
Westreich, Daniel
et al.

Publication Date

2018

DOI

10.1002/bimj.201600140

Peer reviewed

HHS Public Access

Author manuscript

Biom J. Author manuscript; available in PMC 2019 October 01.

Published in final edited form as:

Biom J. 2018 January ; 60(1): 100–114. doi:10.1002/bimj.201600140.

Estimating multiple time-fixed treatment effects using a semi-Bayes semiparametric marginal structural Cox proportional hazards regression model

Stephen R. Cole[1], **Jessie K. Edwards**[1], **Daniel Westreich**[1], **Catherine R. Lesko**[2], **Bryan Lau**[2], **Michael J. Mugavero**[3], **W. Christopher Mathews**[4], **Joseph J. Eron Jr.**[5], **Sander Greenland**[6], **for the CNICS Investigators**

[1]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

[2]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[3]Department of Medicine, School of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA

[4]Department of Medicine, School of Medicine, University of California, San Diego, CA, USA

[5]Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, NC, USA

[6]Departments of Epidemiology and Statistics, UCLA, Los Angeles, CA, USA

Abstract

Marginal structural models for time-fixed treatments fit using inverse-probability weighted estimating equations are increasingly popular. Nonetheless, the resulting effect estimates are subject to finite-sample bias when data are sparse, as is typical for large-sample procedures. Here we propose a semi-Bayes estimation approach which penalizes or shrinks the estimated model parameters to improve finite-sample performance. This approach uses simple symmetric data-augmentation priors. Limited simulation experiments indicate that the proposed approach reduces finite-sample bias and improves confidence-interval coverage when the true values lie within the central “hill” of the prior distribution. We illustrate the approach with data from a nonexperimental study of HIV treatments.

Keywords

bias; causal inference; cohort study; semi-Bayes; semiparametric; survival analysis

Correspondence Stephen R. Cole, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Campus Box 7435, Chapel Hill, NC 27510, USA. cole@unc.edu.

SUPPORTING INFORMATION

Additional Supporting Information including source code to reproduce the results may be found online in the supporting information tab for this article.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

1 | INTRODUCTION

Hernán et al. described a marginal structural Cox proportional hazards model (Cox MSM) for semiparametric survival regression with time-varying treatments (Hernán, Brumback & Robins, 2001). Effects of time-fixed treatments can also be estimated using MSM (Richardson, Kinlaw, MacLehose & Cole, 2015; Sato & Matsuyama, 2003). Conceptually, if we could observe the entire population under each treatment we would know the parameter value. In reality, we only observe a subset of the population, and this subset is restricted such that each patient is observed under only one treatment. Therefore, conditions are required to identify the population effect of a treatment beyond standard regularity conditions (Casella & Berger, 2002, p. 516). Identification is the ability to compute the parameter value that generated the data if the entire (factual but not counterfactual) population were observed. One set of sufficient identification conditions for inverse probability (IP)-weighted fitting of MSMs is: no interference (Hudgens & Halloran, 2008), treatment-version irrelevance (VanderWeele, 2009), no bias due to measurement error (Edwards, Cole & Westreich, 2015), conditional exchangeability (Hernán & Robins, 2006), positivity (Westreich & Cole, 2010), and correct model specification (Platt, Brookhart, Cole, Westreich & Schisterman, 2013). Correct specification is needed for the models used to construct the IP weights, as well as for the structural model used to relate the treatment with the outcome of interest.

In comparative effectiveness research, there is a desire to estimate long-term all-cause and cause-specific survival for a broad set of treatments. When the structural model used for this purpose is of moderate or high dimension relative to the amount of information available, penalized estimation (also known as shrinkage, ridge regression, and partial, empirical, hierarchical, or semi-Bayes estimation; Cox, 1975; Efron & Morris, 1971, 1972, 1973; Good, 1987, 1992; Good & Gaskins, 1971; Morris, 1983) can reduce mean-squared error and thereby improve the accuracy of reported effect estimates from epidemiologic data (Efron & Morris, 1973; Greenland, 1993,1997; Morris, 1983). The Cox proportional hazards model (Cox, 1972) is widely used in comparative effectiveness research, notwithstanding the notable finite-sample bias present in the standard partial-likelihood estimator (Johnson, Tolley, Bryson & Goldman, 1982). The IP-weighted estimator for the parameters of a Cox MSM inherits finite-sample bias from the Cox model, and this bias may be exacerbated by the use of IP weighting (Westreich, Cole, Schisterman & Platt, 2012). Specifically, both ordinary and IP-weighted partial-likelihood fitting of Cox models are subject to finite-sample bias when the number of events per treatment group is not much larger than the number of model covariates. The present paper concentrates on specification of the structural model in the setting of a time-fixed treatment.

In Section 2, we review the Cox MSM and describe a semi-Bayes extension to the model to penalize the resulting estimates in proportion to their distance from a prior mean (usually zero, resulting in shrinkage of the estimates). In Section 3, we present a simulation experiment. In Section 4, we describe our motivating example using data from an ongoing NIH-funded multisite clinical cohort study of HIV-seropositive US adults (Kitahata et al., 2008). We discuss implications of the proposed approach in Section 5.

2 | METHODS

2.1 | Notation

The observed data consist of n individual records $\{Z, X, Y, \Delta\}_i$, where $i = 1,..., n$ indexes patients. We suppress the patient index i below where possible to ease exposition. Uppercase letters represent random variables and lowercase letters represent possible realizations. The vector Z contains the multiple fixed covariates (at study entry). The vector X_j contains the $j = 1,..., \boldsymbol{J}$ fixed (at study entry) treatments; we assume each treatment in X has been centered and scaled so that 0 is a meaningful value and 1 is a meaningful unit difference in values. The observed survival time is $Y = \min(T, C)$, where T is the time from study entry to the event of interest and C is the time from study entry to right-censoring due to administrative end of study or loss to follow-up. If $Y = T$ then the event indicator $\Delta = 1$, else $\Delta = 0$. We assume independent censoring given measured covariates, or formally $f\{T|Z, X\} = f\{T|Z, X, C\}$, where $f(\cdot|\cdot)$ is the conditional density function. We also assume that measurement error is negligible (Hernán & Cole, 2009) and independent between individual processes, including negligible interference (Hudgens & Halloran, 2008).

2.2 | Cox MSM

A possible treatment is denoted by x. The potential survival time indexed by this possible treatment is T^x. A particular potential survival time T^x is factual and thus may be observed if $X = x$, and there is only one version of treatment and no measurement error. When there are multiple versions of treatment (i.e., differences in treatments within a given recognized level of treatment), we must either refine the treatment definition to account for this multiplicity, or assume that treatment version is irrelevant (potential outcomes are independent of version; Cole & Frangakis, 2009; VanderWeele, 2009). A marginal structural Cox proportional hazards regression model is then

$$h_{T^x}(t) = h_0(t)\exp\left\{x\boldsymbol{\beta}\right\}, \quad (1)$$

where $h_0(t)$ is the reference hazard function for $x = 0$, $\boldsymbol{\beta} = (\beta_1,...,\beta_J)$, and $\exp(\beta_j)$ is the causal hazard ratio for a unit change in treatment component x_j. The contribution to the weighted partial-likelihood corresponding to participant i failing at time y_i is

$$L_i(\boldsymbol{\beta}) = \left\{R_i(y_i)\exp\left[X_i\boldsymbol{\beta}\right] / \sum_{k=1}^{n} R_k(y_i)\hat{w}_k(y_i)\exp\left[X_k\boldsymbol{\beta}\right]\right\}^{\hat{w}_i(y_i)},$$

where $R_i(y)$ is an indicator of being in the risk set at time y and $\hat{w}_i(y)$ is the estimated stabilized IP weight. We denote by $\hat{\boldsymbol{\beta}}$ the estimator obtained by maximizing $\prod_{i=1}^{n} L_i(\boldsymbol{\beta})$.

The IP weight is typically the product of treatment and censoring weights, specifically $w(t) = w^X w^C(t)$. In general, the IP weight is a product of component weights, with one

component weight for each conditional exchangeability assumption. The treatment (exposure) weight is

$$w_i^X = \frac{f(X_i)}{f(X_i|Z_i)}.$$

The denominator of this weight corresponds to each participant's probability of receiving their treatment, given the covariates. The censoring weight is

$$w_i^C(t) = \prod_{k=0}^{t} f\left\{C_i(k) = 0 \middle| \bar{C}_i(k-1) = 0, X\right\} / f\left\{C_i(k) = 0 \middle| \bar{C}_j(k-1) = 0, X_i, Z_i\right\},$$

where $\bar{C}(k) = \{C(0), C(1), \ldots, C(k)\}$. The denominator of this weight corresponds to each participant's probability of remaining uncensored up to time t, given the covariates.

Weighting the observed data by these IP weights simulates observing a pseudopopulation in which treatment and censoring are unrelated to the measured covariates Z. Consistency of the estimator $\hat{\boldsymbol{\beta}}$ for treatment effects on the log-hazard follows from a conditional exchangeability assumption, namely the assumption of weak sequential ignorability (no uncontrolled confounding). Specifically, we assume that treatment is mean-independent of each potential outcome given covariates, $E(X|Z) = E(X|Z, T^X)$. An analogous exchangeability assumption must hold for censoring. We also assume correct specification of the models used to estimate the IP weights and correct specification of the hazard model (1).

The standard asymptotic variance estimator for the partial likelihood is not used because $\hat{\boldsymbol{\beta}}$ is a weighted M-estimator (Stefanski & Boos, 2002; Tsiatis, 2006), not a maximum likelihood estimator (MLE). Thus, we instead use the standard asymptotic robust (sandwich) variance estimator, which is conservative in this setting (Hernán et al., 2001).

2.3 | Semi-Bayes Cox MSM

Semi-Bayes (partial-Bayes) methods place prior distributions only on select parameters (Cox, 1975; Greenland, 1992), and can therefore be seen as a Bayes/non-Bayes synthesis (Efron & Morris, 1971, 1972; Good, 1987, 1992; Greenland, 2010). Semi-Bayes methods may be motivated as shrinkage estimators that employ log prior densities as penalties to improve frequency properties of point estimators (such as mean-squared error). These penalties can also improve confidence-interval coverage when the true values of the penalized parameters are in the region of high prior density. We explore these claims below. Unpenalized parameters are those for which prior information is not used, presumably because it is either so weak as to be ignorable, or too difficult or controversial to formulate as a prior distribution.

The fitting method we will employ subtracts a penalty from the IP-weighted log partial likelihood that will shrink the resulting semi-Bayes estimates $\hat{\boldsymbol{\beta}}_{\text{sb}}$ away from the weighted partial-likelihood estimates $\hat{\boldsymbol{\beta}}$ and toward a prior value $\boldsymbol{m} = (m_1, \ldots, m_J)$. A common penalty

is the quadratic penalty, a sum of squared standardized differences between the individual components of $\boldsymbol{\beta}$ and $\boldsymbol{m}$. The resulting penalized log partial likelihood is

$$\log\left\{\boldsymbol{L}(\boldsymbol{\beta})\right\} - (\boldsymbol{\beta} - \boldsymbol{m})' I_p (\boldsymbol{\beta} - \boldsymbol{m})/2,$$

where I_p is the penalty-information matrix, here assumed diagonal with diagonal elements $1/v_j$, and v_j is the prior variance for treatment j. This penalty is equivalent to a log density from a multivariate-normal prior with mean vector $\boldsymbol{m}$ and covariance matrix I_p^{-1}, and corresponds to using independent lognormal priors for the hazard ratios $\exp(\beta_j)$. We maximize this penalized IP-weighted log partial likelihood to obtain the semi-Bayes estimate $\hat{\boldsymbol{\beta}}_{\mathrm{sb}}$, and again use the standard asymptotic robust (sandwich) variance estimator.

We also used log-$F(d, d)$ prior densities (Greenland, 2007a), which are symmetric and heavier tailed than the normal, lighter tailed than t priors, and which approach normality as d increases (Greenland & Mansournia, 2015). These priors have density identical to the standard logistic distribution, with density proportional to $\exp\left(\frac{\beta d}{2}\right)/\{1 + \exp(\boldsymbol{\beta})\}^d$ and hence penalty $\frac{\beta d}{2} - d\log\left\{1 + \exp(\boldsymbol{\beta})\right\}$; for $d = 2$ this prior is identical to the standard logistic distribution, with density proportional to $\exp(\boldsymbol{\beta})/\{1 + \exp(\boldsymbol{\beta})\}^2$. These densities are members of the generalized conjugate family for log-linear models, thus simplifying computation as described below, and correspond to using independent F priors on the hazard ratios $\exp(\beta_j)$.

2.4 | Fitting methods

We use data augmentation to implement the penalization (Greenland, 2003, 2007a, 2007b), which corresponds to fitting the stratified Cox model

$$h_{T^x}(t) = h_{0jk}(t)\exp\left\{x\boldsymbol{\beta}\right\}, \quad j = 0, 1, 2, \ldots, \boldsymbol{J}, \; k = 0, 1,$$

where $j = k = 0$ indicates the actual data hazard function and treatment vector, and $j > 0$, $k = 0, 1$ indexes the augmenting pseudodata encoding the prior information about $\boldsymbol{\beta}$ (prior data). The observed data comprise a stratum indexed here as $k = 0$, $j = 0$; the augmentation data comprise $\boldsymbol{J}$ pairs of strata indexed by $j > 0$, each containing two records indexed by $k = 0, 1$. The pseudodata x_j are set to values which generate a likelihood contribution proportional to the desired prior density for $\boldsymbol{\beta}$.

The "prior data" that generate a normal density for β_j with mean m_j and variance v_j consist of four records. These four records are constructed as two matched-pairs (Greenland, 2003; Greenland & Christensen, 2001; Sullivan & Greenland, 2013). All four records have time $Y = 1$. The first pair has one record with a treated outcome event and one record with an untreated censoring event. The second pair has one record with an untreated outcome event and one record with a treated censoring event. To impose normal priors we employ a scaling factor $s_j = \sqrt{A/\left(2/v_j\right)}$, where $A = 400$ is the prior-data event count and v_j is the prior variance

for β_j. This prior event count A is set arbitrarily large to improve the normal approximation of the data prior (Sullivan & Greenland, 2013). The count A is used as a frequency weight for the prior record, while the IP weights are set to 1 for all four prior-data records. The prior records also include a rescaled treatment indicator $X_j = 1/s_j$, while all other treatment indicators and covariates are set to 0. (While beyond the scope of this paper, if some prior means m_j are nonzero, the prior records must also include an offset variable $f_j = -m_j/s_j$, which is set to 0 in the actual-data records.)

Each set of four records representing a prior data set for β_j are entered in the weighted Cox model as two distinct strata, one for each pair. The correctness of the prior data can thus be checked by running a Cox regression on these four weighted records alone with X_j entered as the only covariate in the model; the resulting estimate of β_j should equal m_j with standard error $v_j^{1/2}$.

The prior data that generate a log-$F(d, d)$ density for β_j are as described above with two alterations (Greenland, 2007a; Sullivan & Greenland 2013). First, we set $A = d/2$. Second, no scaling is required (i.e., $s_j = 1$). The prior data for a hazard ratio thus again consist of four records (in two matched-pairs) with frequency weight $d/2$, which combined represent d events and $2d$ observations, with treatment indicator $X_j = 1$, all other treatment indicators and covariates set to 0. (Again, while beyond the scope of this paper, if there are nonzero m_j, an offset variable is added that is $f_j = -m_j$ in the prior records and 0 in the actual records.)

The correctness of these prior data can again be checked by running a Cox regression on the four weighted records alone; the resulting estimate of β_j should equal m_j with reported standard error of $\sqrt{4/d}$. Further details on data augmentation for Bayesian and penalized estimation can be found in references (Bedrick, Christensen & Johnson, 1996; Cole, Chu & Greenland, 2014; Cole, Chu, Greenland, Hamra & Richardson, 2012; Disacciati, Orsini & Greenland, 2015; Greenland, 2003,2007a, 2007b, 2009a, 2009b; Sullivan & Greenland, 2013).

When using perfectly tied event times (as given in the prior augmentation data) in a Cox model with Breslow's method for ties, the robust variance option in SAS and R employ an adjustment that yields a variance contribution from the prior data that is 1/2 the desired prior variance. Therefore, one must undo this adjustment, which we accomplished using prior data with 1/2 the pair-weights dictated by the above theory to double the variance. Thus, the frequency weights become $A/2$ for the normal prior records and $d/4$ for the log-F prior records. Table A1 presents these "prior data" for the log-$F(2,2)$ application discussed in Section 4.

3 | SIMULATIONS

3.1 | Experimental design

We explored select finite-sample properties of our approach in a setting similar to our example (see Section 4), simulating a cohort of size 10,000 with 200 events with fixed treatment and confounders, and 5000 iterations of the simulation.

We used a binary confounder Z with probability 1/2 for $Z = 1$. The probability of receiving the reference treatment had a marginal probability of 1/5, dependent on Z. If not treated with the referent, one of nine possible alternative treatments had equal probability, each with marginal probability 4/45 ≈ 0.09. The potential outcome under the reference treatment was an exponential random variable T denoting time from treatment assignment to death, with rate parameter dependent on Z. We calculated the potential outcomes under each of the possible alternative treatments by multiplying the reference potential time by the set: 0.2, 0.25, 0.33, 0.5, 1, 2, 3, 4, and 5, corresponding to alternative treatments 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively. We right-censored the combined data at the second percentile of the distribution of T, such that we have 2% events overall. This scenario represents moderate confounding as the binary confounder Z increased the hazard of the outcome 2.72 = exp(1) times and also increased the probability of any nonreferent treatment 2.72 = exp(1) times, but did not influence which of the non-referent treatments were chosen.

We fit three different Cox models for the treatment: unweighted, without the covariate Z (i.e., unadjusted); an IP-weighted Cox model with Z (i.e., a standard marginal structural model); and a penalized and IP-weighted Cox model with Z. While a standard Z-adjusted Cox model could be used in this setting to obtained unbiased covariate-conditional hazard ratios, our intent here is to estimate the covariate-marginal hazard ratios (as in classical direct standardization of rates). The time-fixed IP-weights were estimated using a polytomous logistic regression model, with the 10-level treatment variable as the outcome and the sole covariate Z; these weights were stabilized by the marginal distribution of the treatments as shown in Section 2.2.

We append prior data to the simulation data to implement penalization as described in Section 2, with the priors for each treatment coefficient centered on zero ($m = 0$). The first set of priors for the β_j were normal with 95% of prior probability for each exp(β_j) between either 1/500 and 500, 1/40 and 40, 1/16 and 16, 1/8 and 8, 1/5 and 5, 1/4 and 4, or 1/4 and 2 (corresponding to variances of 10, 3.54, 2, 1.13, 0.67, 0.50, and 0.125). The second set of priors were log-$F(d, d)$ with 95% of prior probability for each exp(β_j) between either 1/648 and 648, 1/39 and 39, 1/15.4 and 15.4, 1/7.15 and 7.15, 1/5 and 5, 1/4 and 4, or 1/2 and 2 (corresponding to $d = 1, 2, 3, 5, 7, 9$, and 33.1).

We calculated percent bias as the ratio of the geometric mean of the 5000 estimated hazard ratios (the antilog of the simulated mean of $\hat{\beta}_j$) divided by the true hazard ratio $R_j = \exp\left(\beta_{j(true)}\right)$, minus 1 times 100. We calculated the "average" standard error of coefficient estimates as the square root of the average of the 5000 robust variance estimates, and the Monte Carlo standard error as the standard deviation of the 5000 estimated log hazard ratios. We calculated the root mean squared error, where the mean-squared error was defined as the squared bias plus the squared MC standard error. We calculated the 95% confidence-interval coverage as the percent of the 5000 Wald-type confidence intervals (using the robust variance) that included the true value; when 94.0% coverage is observed this leads to a 95% confidence interval for our simulation coverage of

$94.0 \pm 1.96\left(6.0 \times \frac{94.0}{5000}\right)^{1/2} = 93.3,\ 94.7$. Finally, we calculated the geometric mean length of the 95% confidence intervals.

3.2 | Experimental results

Results of simulations for percent bias are provided in Table 1. The unadjusted estimates are biased as expected due to confounding. The weighted estimates are decreasingly biased as the effect of the treatment compared to the referent becomes more harmful. This is as expected because when the effect of treatment becomes more harmful, we increase the number of events and thereby reduce finite-sample bias. The sensitivity of the weighted results to finite-sample bias is seen most clearly for the most protective treatment where the true hazard ratio is 0.2, while finite-sample bias does not appear to notably affect the more harmful treatments (e.g., true hazard ratios of 2, 3, 4, and 5).

As expected, the finite-sample bias exaggerates protective effects away from the null (here, downwards) because such finite-sample bias skews the partial likelihood toward infinity. Note that this finite-sample bias (due to the small number of events) is a separate issue from the large-sample confounding bias. Bias is similar for the normal and log-F penalized IP-weighted Cox model estimates, especially when the prior variance was 2/3 or less. The penalized IP-weighted Cox model estimators exhibit little bias when penalization is weak, but show increasing bias toward the null (the prior center) as the penalization becomes stronger, which is especially apparent for the most harmful treatments when finite-sample bias is negligible.

Results of simulations for percent 95% confidence interval coverage are provided in Table 2. Confidence-interval coverage for the unpenalized IP-weighted estimator is 91% or less for true hazard ratios of 0.2 and 0.25, but 94% or more for true hazard ratios of 0.33 or larger, where finite-sample bias becomes unapparent. Both normal and log-F penalized IP-weighted estimators exhibit adequate coverage when penalization is weak, but poor coverage when the penalization becomes strong (i.e., when the prior variance is 1/2 or less), especially for the most protective treatments (those furthest from the center of the prior). The normal priors undercover when there is substantial finite-sample bias (i.e., true hazard ratio of 1/3 or smaller).

Results of simulations for the Monte Carlo standard error and confidence-interval precision are provided in Tables 3 and 4, respectively. The standard error decreases and hence precision increases both as a function of stronger penalization and as the true hazard ratio becomes larger. While the Monte Carlo standard errors are similar for analogous normal and log-F priors, the heavier tails of the log-F priors result in wider confidence intervals and somewhat better coverage (Table 2).

Results for root mean squared error are shown in Figure 1. Both panels show increasing mean-squared error as the treatments become more protective for both unadjusted and the unpenalized IP-weighted estimators. The IP-weighted estimators have notably higher mean-squared error for the most protective treatments. The strongest prior examined (i.e., prior variance of 0.125) exhibits a low root mean squared error until the true hazard ratio was 1/2

or lower, thereafter this prior performs worse than the weaker priors examined. We chose the log-$F(2,2)$ as the primary prior for the application because it exhibited relatively small bias in the range of hazard ratios we expect to see, and had slightly better interval coverage than the analogous (variance 3.54) normal prior, albeit at the expense of slightly increased width.

4 | APPLICATION

4.1 | Description of study and data

We compare 1-year all-cause mortality for 21 initial antiretroviral therapy treatments (Hammer et al., 2006; Thompson et al., 2012) each prescribed to at least 50 patients in the CFAR (Centers for AIDS Research) Network for Integrated Clinical Systems (CNICS) cohort between January 1998 and October 2013. Investigators may request data from the CNICS study by submitting a proposal at https://www.uab.edu/cnics/submit-proposal. We combine the remaining less common treatment plans into an "other" category, so that $J = 22$ in this example. We estimate hazard ratios comparing mortality under each treatment with mortality under a reference treatment consisting of efavirenz, emtricitabine, and tenofovir, which was the most prevalent treatment during the study period.

The study entry coincides with initiation of HIV treatment, which is the time origin for the present pseudoexperiment (i.e., observational study). Eligible patients have no reported history of prior antiretroviral therapy use. While changes in HIV treatment plan may occur over follow-up, here we concentrate on the observational intent-to-treat estimator that ignores changes in treatment (see Cole, Hernán, Margolick, Cohen & Robins, 2005, p. 476).

The outcome of central interest is all-cause mortality, which obviates issues of competing risks (Lau, Cole & Gange, 2009). We have no right-censoring due to dropout because all patients, regardless of clinic follow-up status, were followed for all-cause mortality using the US Social Security Death Index.

In this example the treatment, type of initial HIV therapy, X is fixed at a constant value at therapy initiation. Consequently, confounders Z are also fixed at their value at HIV therapy initiation. The measured covariates are age, sex, race/ethnicity, CD4 count (cells/mm^3), HIV-1 viral load (copies/ml), history of injection drug use, history of male sex with men, history of one or more AIDS diagnoses, and HBV/HCV coinfection, all measured prior to therapy initiation. The stabilized IP weights were estimated using maximum likelihood multinomial logistic regression of treatment regimen on the measured covariates (potential confounders) as the regressors. For each record, the resulting fitted treatment probability for the record enters the denominator of the weight, while the observed proportion in the treatment group of the record (which is the fitted marginal probability of treatment with no covariate adjustment) enters the numerator (Cole & Hernán, 2008; Hernán et al., 2001). Continuous covariates were modeled using restricted quadratic splines with four knots placed at the covariate's 5, 35, 65, and 95 percentiles (Howe et al., 2011).

4.2 | Augmentation data

With the reference treatment consisting of efavirenz, emtricitabine and tenofovir, we assumed for every other treatment independent priors with log-$F(2,2)$ priors for each β_j,

which implies 95% prior probability that the hazard ratio is between 1/39 and 39 and is thus only very weakly informative. In terms of information content under the Cox model, it corresponds to previously observing one exposed and one unexposed event in matched-pairs randomized to each treatment plan. Penalization using other priors examined in the simulation altered results as expected based on the simulations (data not shown).

4.3 | Example results

Table 5 provides the characteristics of the 10,064 patients. Table 6 provides the distributions of the patients and deaths and antiretroviral therapy treatment regimens. Common treatments (with more than 5% of patients) included EFV-FTC-TDF (i.e.,the reference treatment), as well as FTC-TDF-ATV/r, FTC-TDF-DRV/r, and EFV-3TC-ZDV. Notably, 12% of patients were given one of a large number of infrequently used (i.e., less than 50 patients) other treatments.

Table 7 provides the unadjusted, unpenalized IP-weighted, and penalized IP-weighted Cox hazard ratios and 95% confidence intervals for 1-year mortality. For the penalized IP-weighted results, compared to the reference treatment many of the 21 other treatments exhibited elevated hazard ratios for 1-year mortality. Ten regimens exhibited a hazard ratio larger than 1.5, two of which had 95% confidence intervals that excluded the null value of 1 (i.e., other regimen and 3TC-LPV/r-TDF). As expected, penalized estimates were closer to the null with notably tighter confidence intervals. In particular, we obtain a penalized estimate for the treatment ATV/r-3TC-TDF, whereas standard approaches failed to provide an estimate.

5 | DISCUSSION

Both ordinary (Johnson et al., 1982) and IP-weighted (Westreich et al., 2012) partial-likelihood fitting of Cox models are subject to finite-sample bias when the number of events per treatment group is not much larger than the number of model covariates. To reduce this problem, we have presented a semi-Bayes approach to penalize the estimated coefficients in the marginal structural hazard model, paralleling methods for ordinary Cox models (Greenland & Christensen, 2001; Verweij & Van Houwelingen, 1994). The proposed approach was examined in a limited simulation study, and in an example where there were multiple versions of initial treatment.

The log-$F(2,2)$ prior used here is a standard logistic distribution, which corresponds to adding one record with each outcome to a discrete-outcome data set. It thus can be viewed as a generalization of Laplace's law of succession (Laplace, 1814, p. 19; Greenland & Mansournia, 2015) to coefficient estimation (Greenland & Mansournia, 2015). There are other approaches to stabilize estimates from Cox models, including Firth's bias adjustment, which corresponds to use of the Jeffreys invariant prior for penalization (Firth, 1993), and which is implemented in several software packages (Coveney, 2015; Heinze, & Ploner, 2002) including SAS, Stata, and Statistica. The Firth adjustment reduces to the log-$F(1,1)$ prior when only indicators in orthogonal designs are being penalized but otherwise has no straightforward Bayesian interpretation in terms of contextual prior information, and in fact

depends on the data through the observed information matrix (Greenland & Mansournia, 2015).

Under standard first-order asymptotics, the penalized estimators are (such as the MLE) normal, unbiased, and equivalent to the standard (unpenalized) estimator, but this equivalence breaks down at higher orders and hence in finite samples. For example, to second order the MLE is biased away from the null (often severely) while the posterior mode using a Firth penalty (Jeffreys prior) is unbiased (Firth, 1993). In finite samples, as the prior precision $1/v$ or the prior sample size d grows, the penalized estimators become increasingly biased toward the prior mean m.

The extent of penalization varies with the prior, and the substantive impact on estimates of the prior should be explored before inferences are used to set policy or guidelines (Greenland & Mansournia, 2015). In particular, in finite samples where uncertainty about β_j is high, one should choose relatively small values for the penalty tuning parameters (i.e., $1/v$ for the quadratic/normal penalty or d for the log-F penalty). For example, the simulations presented here suggest that independent null-centered log-$F(2,2)$ priors perform well (with respect to minimizing mean-squared error) when treatment effects are in the neighborhood of a hazard ratio of 0.2 to 5, which is typical of much epidemiologic research.

There are two basic approaches to selecting the penalty tuning parameter. The empirical-Bayes approach estimates the parameter from the data (e.g., using marginal maximum likelihood or cross-validation), which requires accounting for this estimation step in subsequent computation of interval estimates and tests. We do not explore this approach, which can be computationally involved (Carlin & Louis, 2010). The second, simpler approach is classical Bayes, in which the parameter is specified on a priori grounds. This can be done by examining the prior probability intervals for the penalized regression coefficients, as implied by various choices for the parameter, then choosing a parameter value that leads to intervals which agree well with results inferred from information outside the current study (Greenland, 1992, 1993, 2003, 2007a, 2007b).

In practice, there is an impression that both of these approaches tend to lead to excessive shrinkage relative to what might be desired based on more careful considerations of possible costs of false negatives and false positives. This concern has led to a “conservative” variation on Bayes, which sets the tuning parameter just large enough to render recognizably extreme coefficient values as highly improbable, without leading to a highly informative concentration of prior probability near the null. For example, in detailed epidemiologic comparative effectiveness research, hazard ratios for study treatments are typically expected to fall in a range of roughly 1/5 to 5 and are rarely found far outside that range, because truly large effects tend to be detected and confirmed well before such sophisticated analyses are initiated. Assuming the outcome is a rare disease, this observation could be incorporated into the penalty by setting $v = 1/2$ or $d = 9$ for the penalty on a coefficient $\boldsymbol{\beta}$ in a Cox model, both of which lead to (1/4, 4) as a 95% prior probability interval for the approximate hazard ratio $\exp(\boldsymbol{\beta})$. When large effects are considered reasonable possibilities, one could reduce d further, for example, $d = 3$ and $d = 1$ correspond approximately to 95% prior probability for $\exp(\beta)$ in (1/16, 16) and (1/40, 40), respectively.

Here, we penalized the parameters in the structural (potential outcome) model. We could have instead or also penalized the parameters in the models used to construct the IP weights. In our example, however, we did not expect the weight models to require such stabilization because the smallest treatment group consisted of 50 people, and there were relatively few parameters in the weight models. Others have explored Bayesian methods for marginal structural modeling that involve priors for the parameters of the weight models (Saarela, Stephens, Moodie & Klein, 2015). While these and other partially Bayesian approaches can improve the frequency properties (calibration) of estimators, fully Bayesian inference is infeasible in typical sparse-data settings due to the breakdown of classical finite-dimensional asymptotic properties of the likelihood (Robins, Hernán & Wasserman, 2015).

For time-fixed treatments, marginal structural Cox models provide consistent estimates of causal effects of treatment under the equally strict conditions required by standard Cox proportional hazards models. Cox MSM analyses estimate marginal treatment effects, which compare hazards averaged (standardized) over the sample covariate distribution, while ordinary Cox models estimate covariate-conditional treatment effects. These estimates can differ even in the absence of confounding because the hazard ratio is not collapsible (Greenland, 1996). As described here, penalization does not alter the target parameter. Therefore, the identification conditions are unaltered by the use of a penalized estimator. Moreover, the penalized estimator remains a consistent estimator of the parameter of interest, as long as the prior information is not a function of sample size. In finite-samples when estimates are shrunk toward the null (i.e., $m = 0$), accompanying confidence intervals will likewise be shifted toward the null and tests based on the confidence limit closest to the null value will be somewhat conservative.

When treatment is time-varying, Cox MSMs maintain consistency under feedback between time-varying treatments and time-varying confounders, while standard time-dependent Cox models do not (Hernán et al., 2001). Future work is needed to assess penalization of a Cox MSM in the setting of time-varying treatments, though theory suggests similar benefits would be found.

Our example is a nonrandomized study analyzed to parallel an experiment in which the initial HIV treatment is randomized and then assigned treatment is analyzed using an intention-to-treat approach, ignoring subsequent regimen nonadherence. Because patients were not randomized to an HIV treatment, we accounted for confounding by measured factors through IP weighting (Hernán et al., 2001; Robins, 1999; Robins, Hernán & Brumback, 2000; Westreich et al., 2012). Nonetheless, uncontrolled confounding may explain some or all of the results.

Measurement errors for select patient characteristics (e.g., age, sex), assigned treatment, and mortality are likely negligible. The impact of measurement errors for other controlled factors (e.g., injection drug use) is unknown, but would be negligible if either the errors are small or if there is relatively modest confounding by these factors (Greenland & Robins, 1985). Selection bias would be negligible because few patients were excluded at entry and no patients were lost to follow-up. The generalizability of our results is uncertain, although recent work suggests that the CNICS cohort may be largely representative of the US HIV

epidemic (Lesko et al., 2016). These caveats would however apply to any standard method as well as all of the analyses we conducted. Within that scope, it appears to us that the results obtained from weakly informative priors are preferable to the usual unpenalized results when we combine frequentist accuracy considerations with background information that the effects under study are probably quite limited.

ACKNOWLEDGMENTS

This research was funded in part by NIH grants R01AI100654, R24AI067039, and P30AI50410. We thank Professors M. Alan Brookhart, Michael G. Hudgens, and James M. Robins for expert advice.

Funding Information

National Institute of Allergy and Infectious Diseases, Grant nos. P30AI50410, R01AI100654, R24AI067039. Gillings Innovation Laboratory award.

APPENDIX

TABLE A1

Log-$F(2,2)$ prior augmentation data example

Record identifler	Regimen	Prior hazard ratio, *hr*	Time	IP-weight	Frequency weight, A[a]	Strata	Regimen 1	…	Regimen *k*	…	Regimen 22	Event indicator	Offset, *f*
1	1	1.0	1	1	0.5	1	1	…	0	…	0	l	0[b]
2	1	1.0	1	1	0.5	1	0	…	0	…	0	0	0
3	1	1.0	1	1	0.5	2	0	…	0	…	0	1	0
4	1	1.0	1	1	0.5	2	1	…	0	…	0	0	0[b]
…	…	…	…	…	…	…	…	…	…	…	…	…	…
85	22	1.0	1	1	0.5	43	0	…	0	…	1	1	0[b]
86	22	1.0	1	1	0.5	43	0	…	0	…	0	0	0
87	22	1.0	1	1	0.5	44	0	…	0	…	0	1	0
88	22	1.0	1	1	0.5	44	0	…	0	…	1	0	0[b]

[a] Noninteger frequency weight is $A = (d/4)$, correcting for the robust variance in SAS or R (see text).

[b] Offset is $f = -\log(hr)$.

REFERENCES

Bedrick EJ, Christensen R, & Johnson W. (1996). A new perspective on priors for generalized linear models. Journal of the Acoustical Society of America, 91, 1450–1460.

Carlin BP, & Louis TA (2010). Bayesian methods for data analysis. New York, NY: Chapman and Hall/CRC.

Casella G, & Berger RL (2002). Statistical inference. Pacific Grove, CA: Duxbury Press.

Cole SR, Chu H, & Greenland S. (2014). Maximum likelihood, profile likelihood, and penalized likelihood: A primer. American Journal of Epidemiology, 179(2), 252–260. [PubMed: 24173548]

Cole SR, Chu H, Greenland S, Hamra G, & Richardson DB (2012). Bayesian posterior distributions without markov chains. American Journal of Epidemiology, 175(5), 368–375. [PubMed: 22306565]

Cole SR, & Frangakis CE (2009). The consistency statement in causal inference: A definition or an assumption? Epidemiology, 20(1), 3–5. [PubMed: 19234395]

Cole SR, & Hernán MA (2008). Constructing inverse probability weights for marginal structural models. American Journal of Epidemiology, 168(6), 656–664. [PubMed: 18682488]

Cole SR, Hernán MA, Margolick JB, Cohen MH, & Robins JM (2005). Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count. American Journal of Epidemiology, 162(5), 471–478. [PubMed: 16076835]

Coveney J. (2015). FIRTHLOGIT: Stata module to calculate bias reduction in logistic regression. Boston, MA: Statistical Software Components Retrieved from https://ideas.repec.org/c7boc/bocode/s456948.html

Cox DR (1972). Regression models and life tables. Journal of the Royal Statistical Society Series B, 34(2), 187–220.

Cox DR (1975). A note on partially Bayes inference and the linear model. Biometrika, 63(3), 651–654.

Disacciati A, Orsini N, & Greenland S. (2015). Approximate Bayesian logistic regression via penalized likelihood by data augmentation. Stata Journal, 15, 712–736.

Edwards JK, Cole SR, & Westreich D. (2015). All your data are always missing: Incorporating bias due to measurement error into the potential outcomes framework. International Journal of Epidemiology, 44(4), 1452–1459. [PubMed: 25921223]

Efron B, & Morris C. (1971). Limiting the risk of Bayes and empirical Bayes estimators: Part 1: The Bayes case. Journal of the Acoustical Society of America, 66, 807–815.

Efron B, & Morris C. (1972). Limiting the risk of Bayes and empirical Bayes estimators: Part 2: The empirical Bayes case. Journal of the Acoustical Society of America, 67, 130–139.

Efron B, & Morris C. (1973). Stein's estimation rule and its competitors: An empirical Bayes approach. Journal of the Acoustical Society of America, 68, 117–130.

Firth D. (1993). Bias reduction of maximum likelihood estimates. Biometrika, 80, 27–38.

Good IJ (1987). Hierarchical Bayesian and empirical Bayesian methods. American Statistician, 41, 92.

Good IJ (1992). The Bayes/non-Bayes compromise: A brief review. Journal of the Acoustical Society of America, 87, 597–606.

Good IJ, & Gaskins RA (1971). Nonparametric roughness penalties for probabilitiy densities. Biometrika, 58, 255–277.

Greenland S. (1992). A semi-Bayes approach to the analysis of correlated multiple associations, with an application to an occupational cancer-mortality study. Statistics in Medicine, 11(2), 219–230. [PubMed: 1579760]

Greenland S. (1993). Methods for epidemiologic analyses of multiple exposures: A review and comparative study of maximum-likelihood, preliminary-testing, and empirical-Bayes regression. Statistics in Medicine, 12(8), 717–736. [PubMed: 8516590]

Greenland S. (1996). Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. Epidemiology, 7(5), 498–501. [PubMed: 8862980]

Greenland S. (1997). Second-stage least squares versus penalized quasi-likelihood for fitting hierarchical models in epidemiologic analyses. Statistics in Medicine, 16(5), 515–526. [PubMed: 9089960]

Greenland S. (2003). Generalized conjugate priors for Bayesian analysis of risk and survival regressions. Biometrics, 59(1), 92–99. [PubMed: 12762445]

Greenland S. (2007a). Bayesian perspectives for epidemiological research. II. Regression analysis. International Journal of Epidemiology, 36(1), 195–202. [PubMed: 17329317]

Greenland S. (2007b). Prior data for non-normal priors. Statistics in Medicine, 26, 3578–3590. [PubMed: 17216667]

Greenland S. (2009a). Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. International Journal of Epidemiology, 38(6), 1662–1673. [PubMed: 19744933]

Greenland S. (2009b). Relaxation penalties and priors for plausible modeling of nonidentified bias sources. Statistical Science, 24(2), 195–210.

Greenland S. (2010). Comment: The need for syncretism in applied statistics. Statistical Science, 25, 158–161.

Greenland S, & Christensen R. (2001). Data augmentation priors for Bayesian and semi-Bayes analyses of conditional-logistic and proportional-hazards regression. Statistics in Medicine, 20(16), 2421–2428. [PubMed: 11512132]

Greenland S, & Mansournia MA (2015). Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. Statistics in Medicine, 34(23), 3133–3143. [PubMed: 26011599]

Greenland S, & Robins JM (1985). Confounding and misclassification. American Journal of Epidemiology, 122(3), 495–506. [PubMed: 4025298]

Hammer SM, Saag MS, Schechter M, Montaner JS, Schooley RT, Jacobsen DM, ... Volberding PA (2006). Treatment for adult HIV infection: 2006 recommendations of the International AIDS Society-USA panel. Topics in HIV Medicine, 14(3), 827–843. [PubMed: 17016878]

Heinze G, & Ploner M. (2002). SAS and SPLUS programs to perform Cox regression without convergence problems. Computer Methods and Programs in Biomedicine, 67(3), 217–223. [PubMed: 11853948]

Hernán MA, Brumback B, & Robins JM (2001). Marginal structural models to estimate the joint causal effect of non-randomized treatments.Journal of the Acoustical Society of America, 96, 440–448.

Hernán MA, & Cole SR (2009). Invited commentary: Causal diagrams and measurement bias. American Journal of Epidemiology, 170(8), 959–962; discussion 963–954. [PubMed: 19755635]

Hernán MA, & Robins JM (2006). Estimating causal effects from epidemiological data. Journal of Epidemiology and Community Health, 60(7), 578–586. [PubMed: 16790829]

Howe CJ, Cole SR, Westreich DJ, Greenland S, Napravnik S, & Eron JJ Jr. (2011). Splines for trend analysis and continuous confounder control. Epidemiology, 22(6), 874–875. [PubMed: 21968779]

Hudgens MG, & Halloran ME (2008). Toward causal inference with interference. Journal of the Acoustical Society of America, 103, 832–842.

Johnson ME, Tolley HD, Bryson MC, & Goldman AS (1982). Covariate analysis of survival data: A small-sample study of Cox's model. Biometrics, 38(3), 685–698. [PubMed: 7171696]

Kitahata MM, Rodriguez B, Haubrich R, Boswell S, Mathews WC, Lederman MM, ... Saag MS (2008). Cohort profile: The Centers for AIDS Research Network of Integrated Clinical Systems. International Journal of Epidemiology, 37(5), 948–955. [PubMed: 18263650]

Laplace PS (1814). A philosophical essay on probabilities (Translated in 1902 by Truscott FW, Emory FW). New York, NY: Wiley.

Lau B, Cole SR, & Gange SJ (2009). Competing risk regression models for epidemiologic data. American Journal of Epidemiology, 170(2), 244–256. [PubMed: 19494242]

Lesko CR, Cole SR, Hall HI, Westreich D, Miller WC, Eron JJ, ... Mugavero MJ (2016); for the CNICS Investigators. Generalizing the effect of antiretroviral therapy on all-cause mortality to the newly HIV-diagnosed population in the United States. International Journal of Epidemiology, 45, 140–150. [PubMed: 26772869]

Morris C. (1983). Parametric empirical Bayes inference: Theory and applications. Journal of the Acoustical Society of America, 78, 47–55.

Platt RW, Brookhart MA, Cole SR, Westreich D, & Schisterman EF (2013). An information criterion for marginal structural models.Statistics in Medicine, 32(8), 1383–1393. [PubMed: 22972662]

Richardson DB, Kinlaw AC, MacLehose RF, & Cole SR (2015). Standardized binomial models for risk or prevalence ratios and differences.International Journal of Epidemiology, 44(5), 1660–1672. [PubMed: 26228585]

Robins JM (1999). Association, causation and marginal structural models. Synthese, 121(1–2), 151–179.

Robins JM, Hernán MA, & Brumback B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology, 11, 550–560. [PubMed: 10955408]

Robins JM, Hernán MA, & Wasserman L. (2015). Discussion of on Bayesian estimation of marginal structural models. Biometrics, 71, 296–299. [PubMed: 25652314]

Saarela O, Stephens DA, Moodie EE, & Klein MB (2015). On Bayesian estimation of marginal structural models. Biometrics, 71, 279–288. [PubMed: 25677103]

Sato T, & Matsuyama Y. (2003). Marginal structural models as a tool for standardization. Epidemiology, 14(6), 680–686. [PubMed: 14569183]

Stefanski LA, & Boos DD (2002). The calculus of M-estimation. American Statistician, 56, 29–38.

Sullivan SG, & Greenland S. (2013). Bayesian regression in SAS software. International Journal of Epidemiology, 42(1), 308–317. [PubMed: 23230299]

Thompson MA, Aberg JA, Hoy JF, Telenti A, Benson C, Cahn P, ... Volberding PA (2012). Antiretroviral treatment of adult HIV infection: 2012 recommendations of the International Antiviral Society-USA panel. Journal of the American Medical Association, 308(4), 387–402. [PubMed: 22820792]

Tsiatis AA (2006). Semiparametric theory and missing data. New York, NY: Springer.

VanderWeele TJ (2009). Concerning the consistency assumption in causal inference. Epidemiology, 20(6), 880–883. [PubMed: 19829187]

Verweij PJ, & Van Houwelingen HC (1994). Penalized likelihood in Cox regression. Statistics in Medicine, 13(23–24), 2427–2436. [PubMed: 7701144]

Westreich D, & Cole SR (2010). Invited commentary: Positivity in practice. American Journal of Epidemiology, 171(6), 674–677; discussion 678–681. [PubMed: 20139125]

Westreich D, Cole SR, Schisterman EF, & Platt RW (2012). A simulation study of finite-sample properties of marginal structural Cox proportional hazards models. Statistics in Medicine, 31, 2098–2109. [PubMed: 22492660]

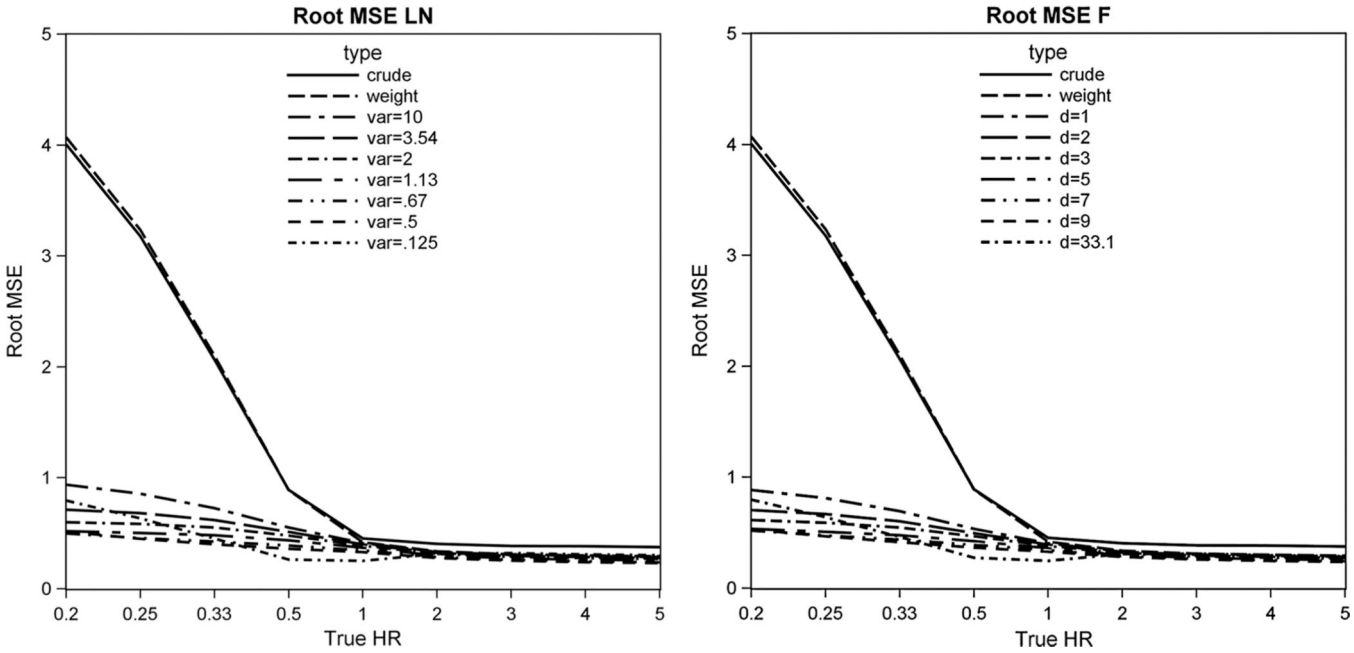

FIGURE 1.

Root mean squared error by true hazard ratio, 5000 simulations of 10,000 participants with 200 events

Note: Left panel displays normal priors, right panel log-*F* priors.

TABLE 1

Percent bias in hazard ratio[a] from 5000 simulations of 10,000 participants with 200 events[b]

			True hazard ratio (expected number of events)								
			0.2 (2.3)	**0.25 (2.8)**	**0.33 (3.8)**	**0.5 (5.7)**	**1 (11.2)**	**2 (22.5)**	**3 (33.4)**	**4 (44.3)**	**5 (54.8)**
Model											
Crude			−66	−46	−15	13	22	27	28	29	29
Weighted			−73	−57	−33	−10	−3	1	1	2	2
Normal priors											
Intervals	Variance										
1/500, 500	10.0		−19	−16	−12	−7	−3	<1	1	2	2
1/40, 40	3.54		−6	−7	−7	−5	−3	<1	<1	1	1
1/16, 16	2.00		4	<1	−2	−3	−3	−1	<1	1	<1
1/8, 8	1.13		18	11	4	−1	−4	−2	−2	−1	−1
1/5, 5	0.67		35	23	12	2	−5	−4	−4	−3	−3
1/2, 4	0.50		46	31	17	5	−6	−6	−6	−5	−5
1/2, 2	0.125		118	84	51	19	−12	−21	−23	−22	−22
Log-$F(d,d)$ priors											
Intervals	Variance	d									
1/648, 648	9.87	1	−15	−13	−10	−6	−3	<1	<1	1	1
1/39, 39	3.29	2	−3	−4	−5	−4	−3	−1	<1	1	1
1/15.4, 15.4	1.87	3	7	3	<1	−2	−4	−1	−1	<1	<1
1/7.15, 7.15	0.98	5	22	14	7	1	−4	−3	−2	−1	−2
1/5, 5	0.66	7	34	23	13	3	−5	−4	−4	−3	−3
¼,4	0.50	9	45	31	18	5	−5	−6	−6	−5	−5
1/2, 2	0.125	33.1	118	85	52	20	−10	−20	−21	−21	−20

[a] 100 × exp(average coefficient estimate)/(True hazard ratio), minus 100.

[b] 20% of participants exposed to the reference treatment (19.4 expected events), 9% to each of these nine comparator treatments.

TABLE 2

Percent coverage of true hazard ratios by Wald-type 95% confidence intervals from 5000 simulations of 10,000 participants with 200 events[a]

			True hazard ratio (expected number of events)								
			0.2 (2.3)	**0.25 (2.8)**	**0.33 (3.8)**	**0.5 (5.7)**	**1 (112)**	**2 (22.5)**	**3 (33.4)**	**4 (44.3)**	**5 (54.8)**
Model											
Unadjusted			82	86	90	91	90	88	88	87	87
Weighted			86	91	94	97	95	95	95	95	95
Normal priors											
Intervals	Variance										
1/500, 500	10.0		86	91	94	96	95	95	95	95	95
1/40, 40	3.54		86	90	93	96	95	95	95	95	95
1/16, 16	2.00		86	90	93	96	95	95	95	95	95
1/8, 8	1.13		88	90	92	95	95	96	95	95	95
1/5, 5	0.67		90	90	92	95	95	96	95	95	95
1/4, 4	0.50		87	90	93	95	95	96	95	95	95
1/2, 2	0.125		0	13	64	94	94	85	80	77	75
Log-$F(d,d)$ priors											
Intervals	Variance	d									
1/648, 648	9.87	1	96	97	96	97	95	96	95	95	95
1/39, 39	3.29	2	96	96	96	97	95	96	95	95	95
1/15.4, 15.4	1.87	3	95	96	96	97	96	96	95	95	95
1/7.15, 7.15	0.98	5	93	94	95	97	96	96	95	96	96
1/5, 5	0.66	7	91	93	95	97	96	96	96	96	96
1/4, 4	0.50	9	88	91	94	97	96	96	96	96	96
1/2, 2	0.125	33.1	7	32	69	94	96	88	85	84	83

[a]20% of participants exposed to the reference treatment (19.4 expected events), 9% to each of these nine comparator treatments.

TABLE 3

Monte Carlo standard error for the log hazard ratio from 5000 simulations of 10,000 participants with 200 events[a]

			True hazard ratio (expected number of events)								
			0.2 (2.3)	**0.25 (2.8)**	**0.33 (3.8)**	**0.5 (5.7)**	**1 (11.2)**	**2 (22.5)**	**3 (33.4)**	**4 (44.3)**	**5 (54.8)**
Model											
Unadjusted			3.86	3.12	2.06	0.88	0.41	0.33	0.30	0.29	0.27
Weighted			3.85	3.12	2.06	0.88	0.42	0.34	0.31	0.30	0.29
Normal priors											
Intervals	Variance										
1/500, 500	10.0		0.91	0.84	0.72	0.55	0.41	0.33	0.31	0.29	0.29
1/40, 40	3.54		0.71	0.68	0.62	0.51	0.40	0.32	0.30	0.29	0.28
1/16, 16	2.00		0.60	0.58	0.55	0.48	0.39	0.31	0.29	0.28	0.27
1/8, 8	1.13		0.49	0.49	0.48	0.44	0.37	0.30	0.27	0.26	0.25
1/5, 5	0.67		0.40	0.41	0.41	0.39	0.34	0.28	0.26	0.25	0.24
1/4, 4	0.50		0.35	0.36	0.37	0.36	0.33	0.27	0.25	0.23	0.23
1/2, 2	0.125		0.16	0.17	0.19	0.20	0.22	0.21	0.19	0.18	0.17
Log-$F(d,d)$ priors											
Intervals	Variance	d									
1/648, 648	9.87	1	0.87	0.80	0.69	0.53	0.40	0.33	0.30	0.29	0.28
1/39, 39	3.29	2	0.70	0.67	0.60	0.49	0.39	0.32	0.29	0.28	0.27
1/15.4, 15.4	1.87	3	0.61	0.59	0.55	0.47	0.38	0.31	0.29	0.27	0.27
1/7.15, 7.15	0.98	5	0.50	0.49	0.47	0.42	0.36	0.30	0.27	0.26	0.25
1/5, 5	0.66	7	0.43	0.43	0.42	0.39	0.34	0.28	0.26	0.25	0.24
1/4, 4	0.50	9	0.38	0.38	0.38	0.36	0.32	0.27	0.25	0.24	0.23
1/2, 2	0.125	33.1	0.17	0.18	0.19	0.21	0.22	0.21	0.20	0.19	0.18

[a]20% of participants exposed to the reference treatment (19.4 expected events), 9% to each of these nine comparator treatments.

TABLE 4

Geometric confidence interval length for the log hazard ratio from 5000 simulations of 10,000 participants with 200 events[a,b]

			True Hazard Ratio (expected number of events)								
			0.2 (2.3)	**0.25 (2.8)**	**0.33 (3.8)**	**0.5 (5.7)**	**1 (112)**	**2 (22.5)**	**3 (33.4)**	**4 (44.3)**	**5 (54.8)**
Model											
Unadjusted			14.96	13.26	10.80	7.50	4.61	3.47	3.31	2.97	2.88
Weighted			15.37	12.63	11.12	7.74	4.80	3.63	3.29	3.12	3.03
Normal priors											
Intervals	Variance										
1/500, 500	10.0		14.92	12.84	10.31	7.35	4.70	3.57	3.24	3.08	2.98
1/40, 40	3.54		10.85	9.99	8.67	6.73	4.54	3.48	3.16	3.00	2.90
1/16, 16	2.00		8.58	8.21	7.48	6.18	4.38	3.38	3.07	2.91	2.82
1/8, 8	1.13		6.56	6.48	6.18	5.46	4.13	3.24	2.94	2.79	2.70
1/5, 5	0.67		5.08	5.12	5.06	4.73	3.84	3.07	2.79	2.64	2.56
1/4, 4	0.50		4.38	4.45	4.46	4.30	3.64	2.96	2.69	2.55	2.46
1/2, 2	0.125		2.41	2.46	2.52	2.60	2.61	2.40	2.23	2.11	2.04
Log-$F(d,d)$ priors											
Intervals	Variance	d									
1/648, 648	9.87	1	24.77	17.16	11.28	7.24	4.63	3.55	3.23	3.06	2.97
1/39, 39	3.29	2	16.31	12.74	9.46	6.67	4.48	3.48	3.17	3.01	2.92
1/15.4, 15.4	1.87	3	12.71	10.53	8.32	6.22	4.34	3.41	3.11	2.95	2.86
1/7.15, 7.15	0.98	5	9.15	8.08	6.86	5.52	4.09	3.28	3.00	2.86	2.77
1/5, 5	0.66	7	7.31	6.69	5.93	5.00	3.88	3.18	2.91	2.77	2.69
1/4, 4	0.50	9	6.19	5.79	5.27	4.60	3.71	3.08	2.84	2.70	2.62
1/2, 2	0.125	33.1	2.85	2.83	2.81	2.77	2.66	2.49	2.37	2.28	2.21

[a] 20% of participants exposed to the reference treatment (19.4 expected events), 9% to each of these nine comparator treatments.

[b] Geometric length is the antilog of the mean length of 5000 log hazard ratio confidence intervals.

TABLE 5

Characteristics of 10,064 HIV-seropositive CNICS patients initiating ART, 1998–2013

Characteristics	Number	Percentage
Female	1743	17%
Age (years)[a]	39	32; 46
Hispanic ethnicity	1249	12%
African–American race	3769	38%
Injection drug user, current or past use	1414	14%
Men who have sex with men	6276	62%
CD4 cell count (cells/ml^3)[a]	263	1G6; 417
HIV-1 viral load (copies/ml)[a]	41,853	5834; 145,754
History of one or more AIDS diagnoses	2283	23%
HCV coinfection	1211	12%
HBV coinfection	6G7	6%

Note: ART, antiretroviral therapy.

[a]Median, quartiles.

TABLE 6

Distribution of 10,064 HIV-seropositive CNICS patients and deaths by initial ART regimens, 1998–2013

	Patients		Deaths, 1 year	
ART regimen	**Number**	**Percentage**	**Number**	**Percentage**
ABC 3TC ATV/r	154	2	8	4
ABC EFV 3TC	149	1	5	2
ABC EFV 3TC ZDV	108	1	6	3
ABC 3TC LRV/r	59	1	6	3
ABC 3TC LPV/r ZDV	65	1	3	1
ABC 3TC ZDV	271	3	7	3
ATV/r FTC TDF	1391	14	23	11
ATV/r 3TC TDF	53	1	0	0
DRV/r FTC TDF	680	7	12	6
EFV FTC TDF	3313	33	33	16
EFV 3TC d4T	165	2	7	3
EFV 3TC TDF	162	2	5	3
EFV 3TC ZDV	654	7	11	5
FTC FPV/r TDF	88	1	3	1
FTC LPV/r TDF	295	3	10	5
FTC NVP TDF	65	1	1	1
FTC RAL TDF	397	4	4	2
FTC RPV TDF	256	3	1	1
3TC LPV/r TDF	133	1	10	5
3TC LPV/r ZDV	258	3	5	2
3TC NVP ZDV	112	1	2	1
Other	1216	12	39	19
Total	10,066	100	196	100

Note: ART, antiretroviral therapy.

TABLE 7

Hazard ratios for 1-year all-cause mortality by initial ART regimen, 1998–2013

					Penalized[b]	
ART regimen	**Crude hazard ratios**	**95% CI**	**IP-weighted[a] Hazard ratios**	**95% CI**	**IP-weighted[a] Hazard ratios**	**95% CI**
EFV FTC TDF	1	–	1	–	1	–
FTC RPV TDF	0.38	0.05, 2.81	0.31	0.04, 2.57	0.38	0.08, 1.89
FTC RAL TDF	1.07	0.38, 3.03	0.71	0.22, 2.32	0.71	0.24, 2.11
DRV/r FTC TDF	1.66	0.85, 3.22	1.31	0.63, 2.72	1.27	0.62, 2.57
ABC EFV 3TC ZDV	4.09	1.69, 9.88	1.61	0.61,4.26	1.50	0.58, 3.86
ATV/r FTC TDF	1.66	0.97, 2.83	1.07	0.61, 1.89	1.04	0.61, 1.78
EFV 3TC ZDV	1.46	0.73, 2.90	0.66	0.30, 1.45	0.65	0.31, 1.37
3TC NVP ZDV	1.99	0.48, 8.30	1.07	0.22, 5.27	1.03	0.24, 4.43
ATV/r 3TC TDF	NA	NA	NA	NA	0.39	0.05, 2.83
ABC EFV 3TC	2.70	1.05, 6.94	1.83	0.59, 5.64	1.72	0.57, 5.15
FTC LPV/r TDF	3.20	1.57, 6.54	2.32	1.01,5.35	2.22	0.98, 5.02
ABC 3TC ZDV	2.23	0.98, 5.05	1.54	0.62, 3.82	1.48	0.61, 3.55
ABC ATV/r 3TC	4.49	2.06, 9.76	1.57	0.63, 3.92	1.49	0.61, 3.60
3TC LPV/r ZDV	1.70	0.66, 4.36	1.77	0.59, 5.29	1.68	0.58,4.91
EFV 3TC TDF	2.84	1.10, 7.29	1.52	0.55, 4.25	1.44	0.54, 3.89
ABC 3TC LPV/r ZDV	3.64	1.11, 11.9	0.89	0.23, 3.47	0.88	0.27, 2.94
Other	2.95	1.85, 4.71	1.84	1.13,3.00	1.78	1.12, 2.82
3TC LPV/r TDF	6.92	3.38, 14.2	3.50	1.25,9.82	3.24	1.15, 9.15
FTC NVP TDF	1.47	0.20, 10.8	2.21	0.31, 15.8	1.93	0.28, 13.3
ABC 3TC LRV/r	8.47	3.52, 20.4	2.53	0.78, 8.28	2.26	0.70, 7.30
EFV 3TC d4T	3.87	1.70, 8.81	2.05	0.80, 5.27	1.93	0.76, 4.87
FTC FPV/r TDF	2.77	0.85, 9.06	2.40	0.67, 8.62	2.18	0.61,7.83

Note: ART, antiretroviral therapy; CI, confidence interval.

[a] Inverse-probability weights for characteristics shown in Table 1, with continuous variables fit using restricted quadratic splines.

[b] Penalized with log-$F(2,2)$ data augmented prior hazard ratios of 1 with 95% prior mass between 1/39 and 39.