

RESEARCH

Open Access



The impact of violation of the proportional hazards assumption on the discrimination of the Cox proportional hazards model

Peter C. Austin^{1,2,3*} and Daniele Giardiello⁴

Abstract

Background The Cox proportional hazards regression model is frequently used to estimate an individual's probability of experiencing an outcome within a specified prediction horizon. A key assumption of this model is that of proportional hazards. An important component of validating a prediction model is assessing its discrimination. Discrimination refers to the ability of predicted risk to separate those who do and do not experience the event. The impact of violation of the proportional hazards assumption on the discrimination of risk estimates obtained from a Cox model has not been examined.

Methods We used Monte Carlo simulations to assess the impact of the magnitude of the violation of the proportional hazards assumption on the discrimination of a Cox model as assessed using the time-varying area under the curve and on predictive accuracy as assessed using the time-varying index of predictive accuracy.

Results Compared to settings in which the proportional hazards assumption was satisfied, discrimination and predictive accuracy decreased in settings in which the log-hazard ratio was positively associated with time. Conversely, compared to settings in which the proportional hazards assumption was satisfied, discrimination and predictive accuracy increased in settings in which the log-hazard ratio was negatively associated with time. Compared with the use of a Cox regression model, the use of accelerated failure time parametric survival models, Royston and Parmar's spline-based parametric survival models, and generalized linear models using pseudo-observations did not result in estimates with improved discrimination or predictive accuracy in settings in which the proportional hazards assumption was violated.

Conclusions Violation of the proportional hazards assumption had an effect on the discrimination of predictions obtained using a Cox regression model.

Keywords Cox regression, Proportional hazards, Discrimination, Model validation, Monte carlo simulations

*Correspondence:

Peter C. Austin

peter.austin@ices.on.ca

¹ICES V106, 2075 Bayview Avenue, Toronto, ON M4N 3M5, Canada

²Institute of Health Policy, Management and Evaluation, University of Toronto, Ontario, Canada

³Sunnybrook Research Institute, Toronto, Ontario, Canada

⁴Bicocca Bioinformatics Biostatistics and Bioimaging B4 Center, School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

The Cox model characterizes the relationship between covariates and the hazard function, which represents the instantaneous rate of event occurrence at a given time, conditional on survival up to that time. A key assumption of the Cox regression model is that of proportional hazards: the ratio of the hazard function for any two subjects is constant over time and is a function only of their covariates and the regression coefficients. While the Cox model allows one to estimate the magnitude of the association between covariates and the hazard function, it is also frequently used to estimate an individual's absolute risk of experiencing an outcome (e.g., death) within a specified prediction horizon (e.g., 10 years) in settings in which the length of follow-up varies across individuals and where some individuals can be subject to censoring prior to the prediction horizon. This can be done by combining Breslow's estimate of the cumulative hazard function and with the linear predictor comprised of the values of the predictor variables combined with the estimated regression coefficients [1].

Assessing calibration and discrimination are two important aspects of validating the performance of clinical prediction models [2–4]. Calibration refers to the agreement between predicted and observed risks while discrimination refers to the ability of predicted risk to discriminate between those with and without the event. A recent study examined the impact of the violation of the proportional hazards assumption on the calibration of a Cox proportional hazards regression model. It concluded that, in general, violation of the proportional hazards assumption had a negligible impact on calibration [5]. Royston and Altman suggested that, even if the proportional hazards assumption is violated, a Cox model 'may still provide good discrimination' [6]. However, there is a paucity of formal research on the effect of the violation of the proportional hazards assumption on the discrimination of clinical prediction models developed using the Cox regression model.

The objective of the current paper was to examine the impact of the violation of the proportional hazards assumption on the discrimination of predicted event probabilities within a specified prediction horizon generated from a Cox proportional hazards regression model. The paper is structured as follows: In Sect. 2, we describe the design of a series of Monte Carlo simulations that were used to examine this issue. In Sect. 3, we report the results of these simulations. In Sect. 4, we provide a sensitivity analysis in which we assess the impact of varying the prevalence of a binary variable for which the proportional hazards assumption is violated. Finally, in Sect. 5, we summarize our findings and place them in the context of the existing literature.

Monte Carlo simulation methods

We used Monte Carlo simulations to examine the impact of violation of the proportional hazards assumption on the discrimination of predicted event probabilities within a specified prediction horizon generated from a Cox proportional hazards model. The data-generating process used was similar to that used in a previous study that examined the impact of violation of the proportional hazards assumption on the calibration of the Cox model [5]. Motivated by that prior study, we compared the discrimination of predicted event probabilities within a specified prediction horizon generated from a Cox regression model under violation of the proportional hazards assumption with alternative methods for estimating risk with time-to-event outcomes: a Cox model that stratified on the variable for which the proportional hazard assumption was violated (when the variable for which the assumption was violated was binary), parametric accelerated failure time (AFT) survival models, Royston and Parmar's spline-based flexible parametric survival models, and generalized linear models based on pseudo-observations.

Data-generating process

Our data-generating process was similar to that in a previous study that examined the impact of violation of the proportional hazards assumption on the calibration of the Cox regression model [5].

For each subject in a sample of size 2,000, we generated two correlated baseline covariates: a continuous variable (X_1) and a binary variable (X_2). To do so, we first generated random variates from a bivariate normal distribution with mean (0,0) and variance-covariance

matrix $\begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$. The first variable, X_1 , was retained as a continuous variable, while the second variable, X_2 , was dichotomized such that it was set equal to 1 if it was less than zero and set equal to 0 otherwise.

Using previously-described methods, we generated time-to-event outcomes from the following hazards model: $\log(h(t)) = \log(h_0(t)) + (\beta_0 + \beta_1 t)X_1 + \beta_2 X_2$, where t denotes time, $h(t)$ denotes the hazard function and $h_0(t)$ denotes the baseline hazard function [7]. Under this data-generating process, the proportional hazards assumption was violated for X_1 , the continuous baseline covariate, since the regression coefficient for X_1 is a linear function of time. When simulating time-to-event outcomes, we assumed that the maximum observed event time was 1,825 days (i.e., 5 years). We allowed the value of β_1 to vary from -0.0008 to 0.0008 in increments of 0.0001 , thereby varying the magnitude of the violation of the proportional hazards assumption. For a given value of β_1 , we determined the value of β_0 so that the average

regression coefficient across the 1,825 days would be equal to $\log(2)$ (therefore the average hazard ratio for a one standard deviation increase in X_1 across the 1,825 days would be equal to 2). We thus examined 17 different scenarios, with one scenario ($\beta_1 = 0$) denoting a scenario in which the proportional hazards assumption was satisfied. Increasing values of $|\beta_1|$ denote stronger violations of the proportional hazards assumption. The regression coefficient for the binary variable (β_2) was set to equal to $\log(2)$. This procedure was repeated 1,000 times so that for each simulation scenario we created 1,000 simulated datasets. The decision to set the average regression coefficient across the 1,825 days to $\log(2)$ was made for a pragmatic reason. When combined with different values of β_1 , this allowed for the hazard ratio to display minor, moderate, or strong variation over the 1,825 days of follow-up, while resulting in time-specific hazard ratios that would be plausible in a variety of settings.

We then modified the above data-generating process so that the proportional hazards assumption was violated for the binary variable and satisfied for the continuous variable. We thus had a set of Monte Carlo simulations in which we allowed two factors to vary: (i) whether the variable for which the proportional hazards assumption

was violated was binary or continuous; (ii) the magnitude of the violation of the proportional hazards assumption. The first factor took on two values while the second factor took on 17 values. In each of the 34 scenarios we simulated 1,000 datasets.

The time-varying hazard ratios in each of the 17 scenarios are reported in Fig. 1. The left panel describes the time-varying hazard ratios when the proportional hazards assumption was violated for the binary variable, while the right panel describes the time-varying hazard ratios when the proportional hazards assumption was violated for the continuous variable. The figure illustrates that we considered scenarios with mild violation of the proportional hazards assumption to scenarios with strong violation of the proportional hazards assumption. Furthermore, we considered scenarios in which the hazard ratio decreased over time ($\beta_1 < 0$) and scenarios in which the hazard ratio increased over time ($\beta_1 > 0$).

Analyses in the simulated datasets

Each simulated dataset was split into a derivation sample and a validation sample of equal size. In each derivation sample we regressed the hazard of the outcome on the two baseline covariates using the following Cox regression

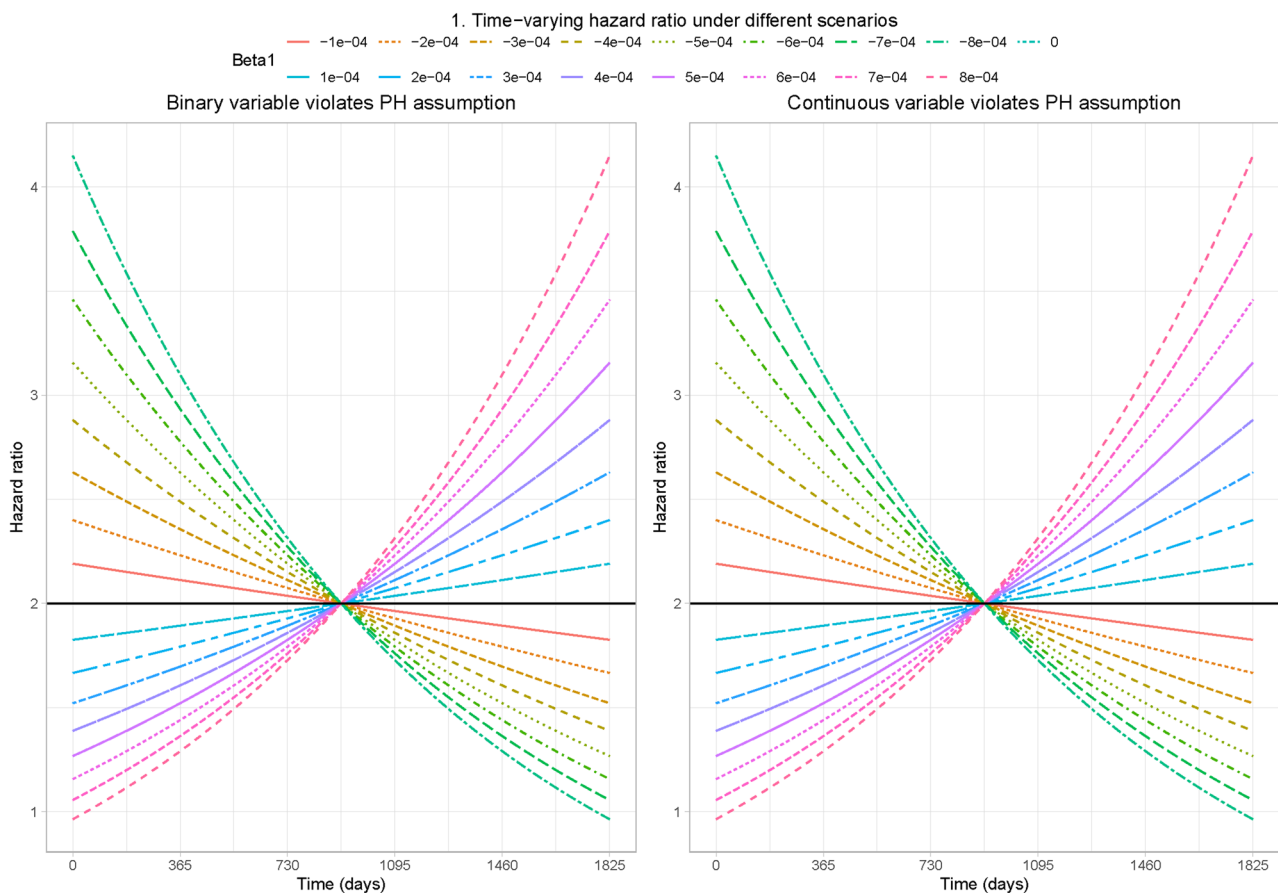


Fig. 1 Time-varying hazard ratio under different scenarios

model: $\log(h(t)) = \log(h_0(t)) + \beta_1 X_1 + \beta_2 X_2$ (i.e., we fit a misspecified model in which we ignored potential violation of proportional hazards assumption).

The fitted model was then applied to the validation sample and the linear predictor was computed for each subject in the validation sample. Harrell's concordance statistic (which he referred to as an index of discrimination of a prognostic model) was computed in the validation sample [8]. We also computed Uno's c-statistic [9].

We then obtained a predicted probability of the outcome within 1, 2, 3, 4, and 5 years for each subject in the validation sample. Breslow's estimate of the cumulative baseline hazard function was used to compute each individual's probability of experiencing the outcome within each of the five prediction horizons. Discrimination of the model in the validation sample was assessed at each of the five time points using the time-varying area under the receiver operating characteristic (ROC) curve, which we refer to as the time-varying AUC [10]. We also computed the Index of Prediction Accuracy (IPA), which is equivalent to the scaled Brier score, at each of the five prediction horizons (strictly speaking the IPA is not a measure of only discrimination, but combines

discrimination and calibration; we include it here for the sake of completeness) [11].

We also examined the performance of three other methods for obtaining estimates of risk at specified durations of time: (i) an AFT parametric survival model with a gamma distribution; (ii) Royston and Parmar's spline-based flexible parametric survival models using splines with four degrees of freedom; (iii) generalized linear models that used pseudo-observations [12, 13]. With the latter approach, we computed pseudo-observations at each of the five prediction horizons and regressed the pseudo-observations at a given prediction horizon on the two baseline covariates using a generalized linear model with a normal distribution and a bounded logit link function (we then repeated this method using a bounded complementary log-log link function) [14]. We then applied the fitted linear model to the validation sample to obtain the probability of the occurrence of the outcome at the given prediction horizon. For the AFT survival model, the following model was fit: $\log(T) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \sigma \epsilon$, where T denotes survival time, σ is a scale parameter, and ϵ is a random error term whose distribution is determined by the

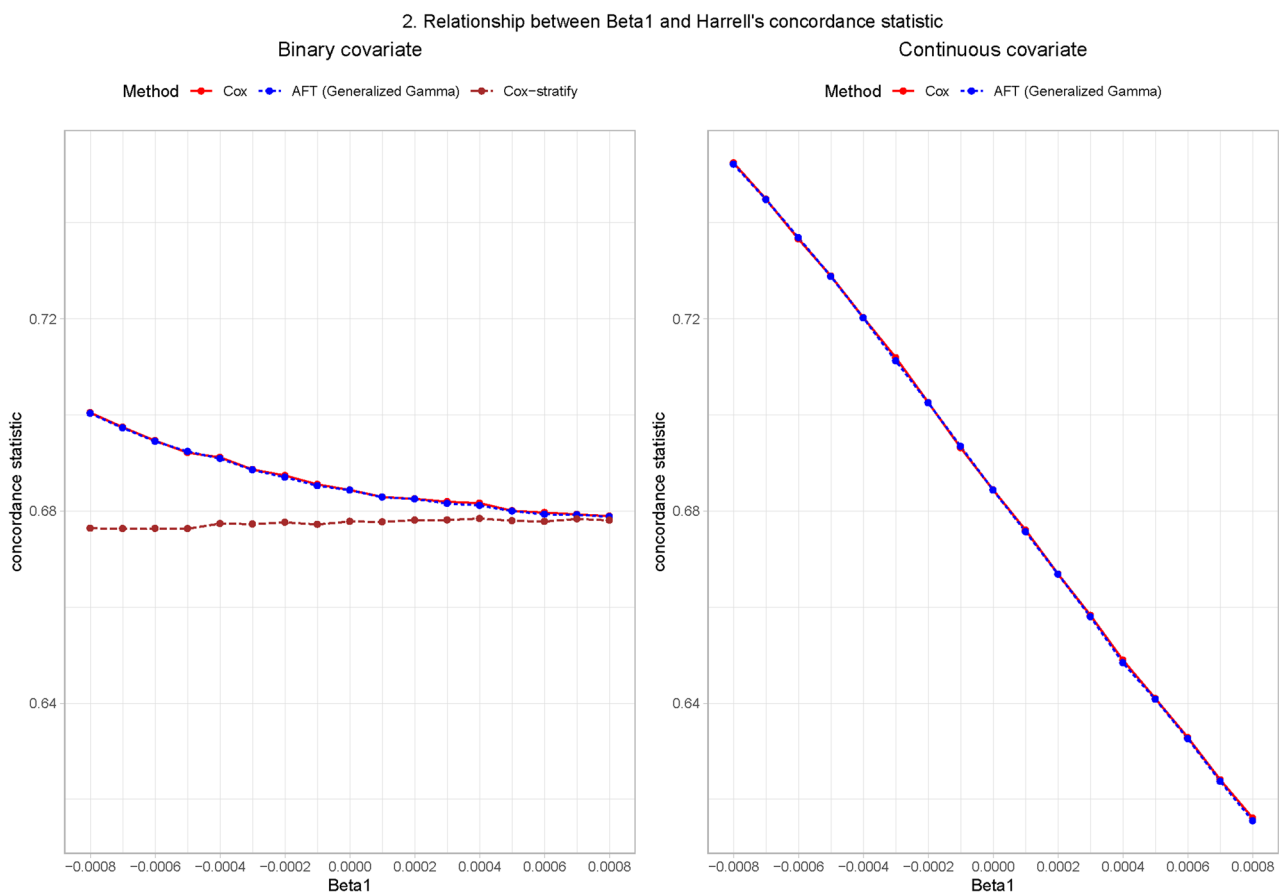


Fig. 2 Relationship between beta1 and harrell's concordance statistic

gamma distribution. For Royston and Parmar’s spline-based parametric survival model, the following model was fit: $\log(H(t)) = s(x; \gamma) + \beta_1 X_1 + \beta_2 X_2$, where $H(t)$ is the cumulative hazard function and $s(x; \gamma)$ is a k-knot natural cubic spline function of log-time that is used to model the baseline cumulative hazards function (in our application we used a splines with 4 degrees of freedom). For the generalized linear models using pseudo-values, the linear predictor was modeled as $\beta_0 + \beta_1 X_1 + \beta_2 X_2$.

In those settings in which the proportional hazards assumption was violated for the binary variable, we also fit a Cox regression that stratified on the binary variable, thereby allowing a separate baseline hazard function for each of the two levels of the binary variable. This stratified model had a single predictor variable: the continuous baseline variable.

The mean of each performance measure was computed across the 1,000 simulated samples for each simulation scenario. Harrell’s and Uno’s concordance statistics were only computed for the Cox regression model and the gamma AFT survival model, as these concordance statistics require a linear predictor for each individual, rather than an estimated risk at specific prediction windows.

Software

The simulations were conducted using the R statistical programming language (version 3.6.3) [15]. Time-to-event outcomes were simulated using the `simsurv` function from the `simsurv` package (version 1.0.0). The Cox model was fit using the `coxph` function from the `survival` package (version 3.2–11). Harrell’s and Uno’s c-statistics were computed using the `concordance` function from the `survival` package. Predicted risk from the fitted Cox model was estimated using the `predictSurvProb` function from the `pec` package (version 2019.11.03). The generalized gamma AFT model was fit using the `flexsurvreg` function from the `flexsurv` package (version 2.3). Royston and Parmar’s spline-based parametric survival model was fit using the `stpm2` function from the `rstpm2` package (version 1.5.2). Pseudo-observations were computed using the `pseudosurv` function from the `pseudo` package (version 1.4.3). The generalized linear model with the pseudo-observations was fit using the `glm` function using the bounded logit link function `blogit` from the `survival` package. The bounded complementary log-log link function was implemented using the `cloglog` function from the `survival` package. Time-vary AUC and IPA were computed using the `Score` function from the `riskRegression`

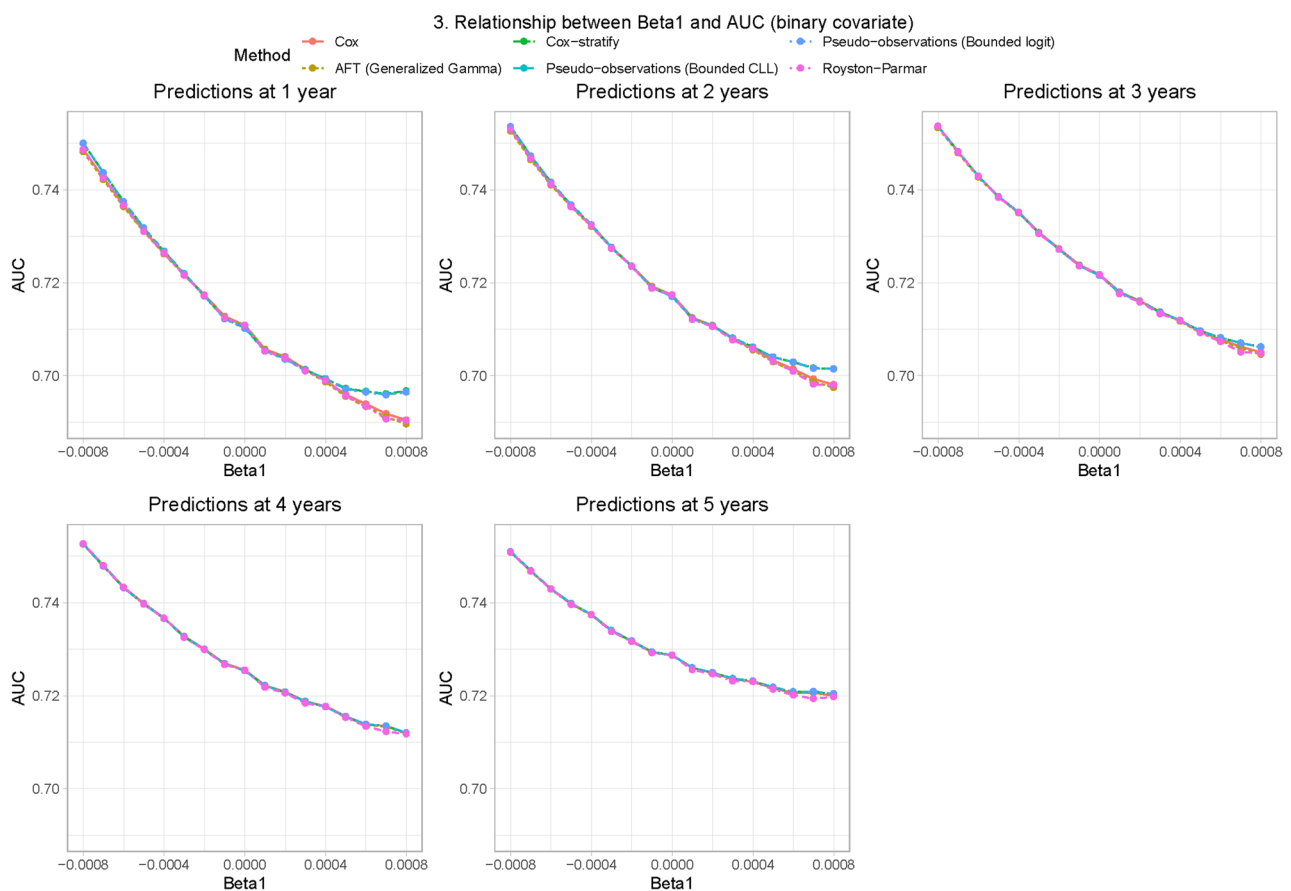


Fig. 3 Relationship between Beta1 and AUC (binary covariate)

package (version 2020.12.08). Simulation results were summarized using the `simsum` function from the `rsumsum` package (version 0.13.0).

Monte Carlo simulation results

We report our results separately for the scenarios when the variable for which the proportional hazards assumption was violated was binary and when it was continuous. When examining the figures used to summarize our findings, recall that $\beta_1 = 0$ denotes a scenario in which the proportional hazards assumption was satisfied.

Proportional hazards assumption violated for a binary variable

The relation between β_1 and Harrell's concordance index is described in the left panel of Fig. 2. For the unstratified Cox model and for the AFT survival model, compared to when the proportional hazards assumption was satisfied ($\beta_1 = 0$), increasingly larger positive values of β_1 resulted in negligible decreases in the concordance index. Conversely, compared to when the proportional hazards assumption was satisfied ($\beta_1 = 0$), decreasing values of β_1 were associated with modest increases in the concordance index. Thus, the largest concordance index was

observed when the time-varying hazard ratio was initially very large and then decreased over time, while the smallest concordance index was observed when the time-varying hazard ratio was initially low and then increased over time. For the stratified Cox model, violation of the proportional hazards assumption had no impact on concordance. Essentially identical results were obtained for Uno's c-statistic (results not shown). The left panel of Figure S1 in the supplemental online material reports the relationship between β_1 and the standard deviation of the computed concordance statistics across the 1,000 simulation replicates. In general, the concordance statistics derived from the two Cox-based models displayed greater variability across simulation replicates than did the concordance statistics derived from the AFT survival model.

The relation between β_1 and time-varying AUC is reported in Fig. 3, which consists of five panels, one for each of the five prediction horizons. At each prediction horizon, the relation between β_1 and time-varying AUC was approximately linear, with AUC decreasing as β_1 increased. The magnitude of the change in AUC with increasing β_1 decreased as the prediction horizon increased (e.g., at the 1-year prediction horizon the

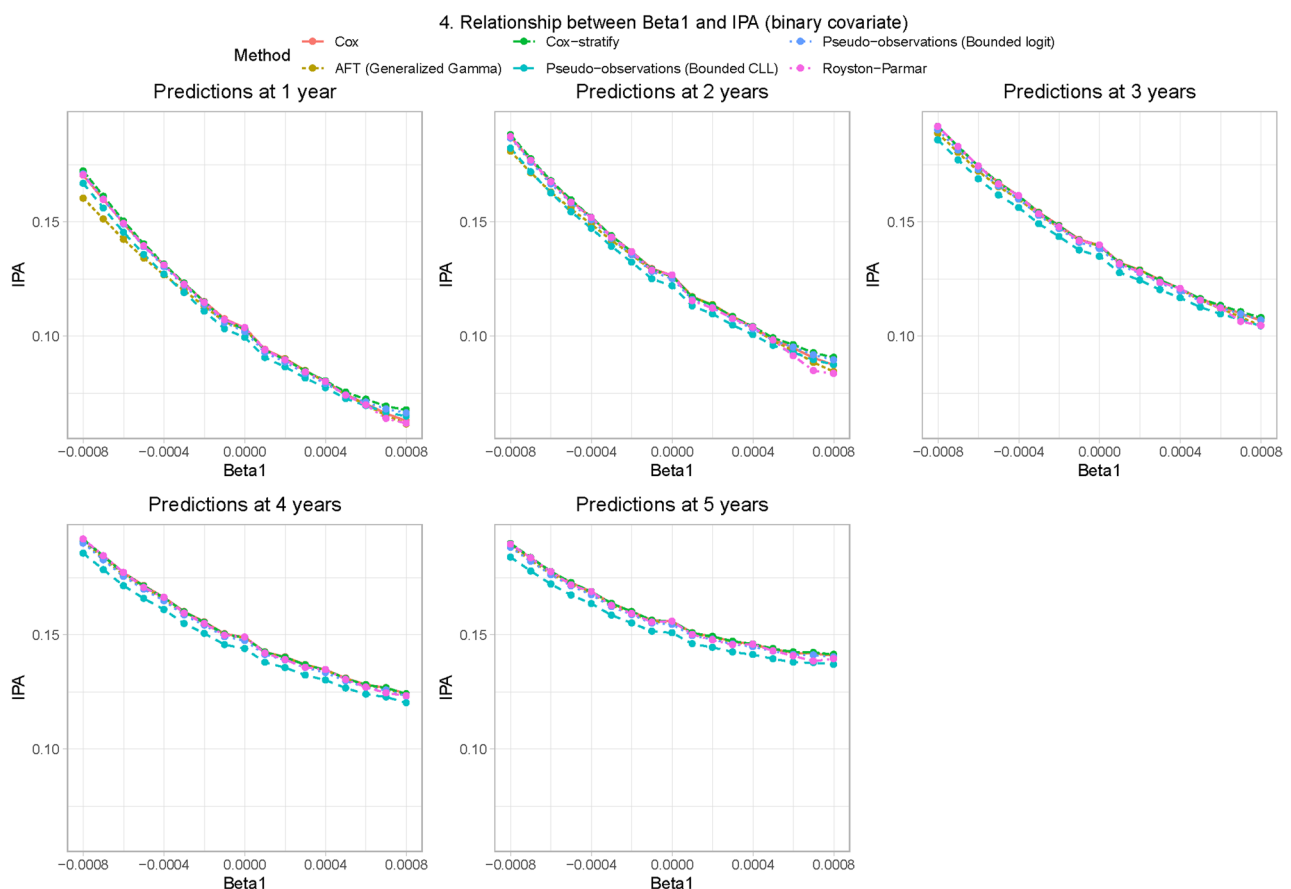


Fig. 4 Relationship between Beta1 and IPA (binary covariate)

change in AUC as β_1 increased from -0.0008 to 0.0008 was approximately 0.06, while at the 5-year prediction horizon the change was approximately 0.03). At each prediction horizon, the largest AUC was observed when $\beta_1 = -0.0008$, the scenario in which there was a very strong violation of the proportional hazards assumption and where the time-varying hazard ratio was initially very large and then decreased over time. The Monte Carlo standard errors of the mean AUC are described in Figure S2 in the supplemental online material [16]. For most scenarios, prediction horizons, and statistical models, the Monte Carlo standard errors lay between 0.0005 and 0.0008. When interpreting these Monte Carlo standard errors in the context of the magnitude of the estimated AUCs in Fig. 3, this would suggest that the mean AUCs were estimated with high precision. Figure S3 in the supplemental online material reports the relationship between β_1 and the standard deviation of the computed AUCs across the 1,000 simulation replicates. There is one panel for each of the five prediction horizons. The computed AUCs from five of the six models displayed a similar variability across simulation replicates. The one exception was that, compared to the other five methods, the AUCs from the Royston-Parmer model displayed

greater variability across simulation replicates in several of the scenarios.

The relation between β_1 and the time-varying IPA is reported in Fig. 4, which consists of five panels, one for each of the five prediction horizons. The relation between β_1 and the time-varying IPA was very similar to that observed for the time-varying AUC. Figure S4 in the supplemental online material reports the relationship between β_1 and the standard deviation of the computed IPAs across the 1,000 simulation replicates. The computed IPAs from five of the six models displayed a similar variability across simulation replicates. The one exception was that, compared to the other five methods, the IPAs from the Royston-Parmer model displayed greater variability across simulation replicates in several of the scenarios.

Proportional hazards assumption violated for a continuous variable

The relation between β_1 and Harrell’s concordance index was approximately linear (right panel of Fig. 2). The largest concordance index was observed when the time-varying hazard ratio was initially very large and then decreased over time ($\beta_1 = -0.0008$), while the smallest

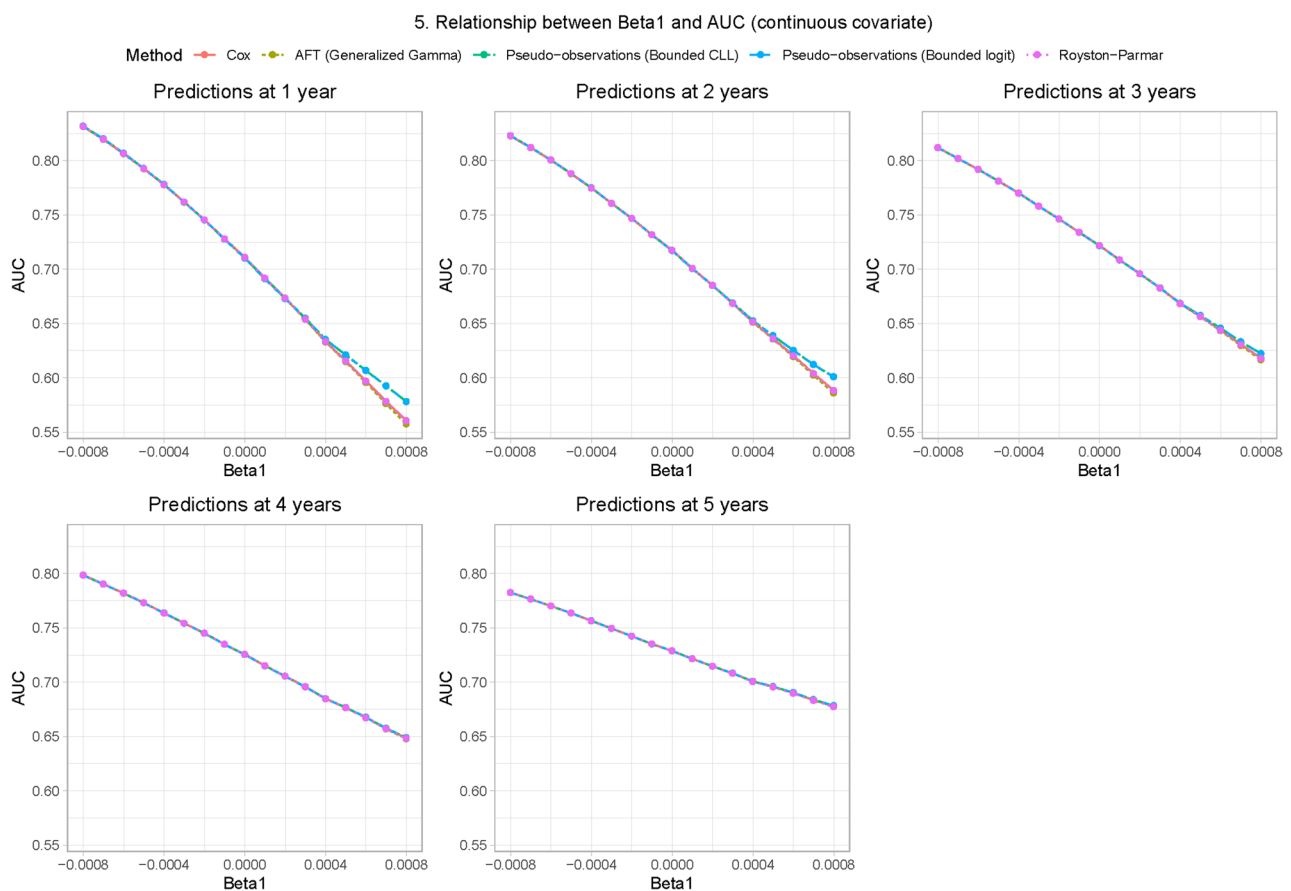


Fig. 5 Relationship between Beta1 and AUC (continuous covariate)

concordance index was observed when the time-varying hazard ratio was initially low and then increased over time ($\beta_1 = 0.0008$). Essentially identical results were obtained for Uno's c-statistic (results not shown). The right panel of Figure S1 in the supplemental online material reports the relationship between β_1 and the standard deviation of the computed concordance statistics across the 1,000 simulation replicates. In several scenarios, the concordance statistic derived from the Cox regression model displayed greater variability across simulation replicates than did the concordance statistic derived from the AFT survival model.

At each prediction horizon, the relation between β_1 and time-varying AUC was approximately linear, with AUC decreasing as β_1 increased (Fig. 5). The magnitude of the change in AUC with increasing β_1 decreased as the prediction horizon increased. At each prediction horizon, the largest AUC was observed when $\beta_1 = -0.0008$, the scenario in which there was a very strong violation of the proportional hazards assumption, where the time-varying hazard ratio was initially very large and then decreased over time. The Monte Carlo standard errors of the mean AUC are described in Figure S5 in the supplemental online material. For most scenarios, prediction

horizons, and statistical models, the Monte Carlo standard errors lay between 0.00045 and 0.0008, indicating that the mean AUCs were estimated with great precision. Figure S6 in the supplemental online material reports the relationship between β_1 and the standard deviation of the computed AUCs across the 1,000 simulation replicates. In general, the computed AUCs from four of the five models displayed similar variability across simulation replicates. The one exception was that, compared to the other four methods, the AUCs from the Royston-Parma model displayed greater variability across simulation replicates in several of the scenarios.

The relation between β_1 and the time-varying IPA is reported in Fig. 6. The relation between β_1 and the time-varying IPA was very similar to that observed for time-varying AUC. Note that at the 1-year prediction horizon with a very strong violation of the proportional hazards assumption ($\beta_1 = 0.0008$), the IPA was less than zero for most prediction methods, indicating that the fitted model produces less accurate predictions than the null model that produces the same estimated risk for all individuals. Figure S7 in the supplemental online material reports the relationship between β_1 and the standard deviation of the computed IPAs across the 1,000 simulation replicates.

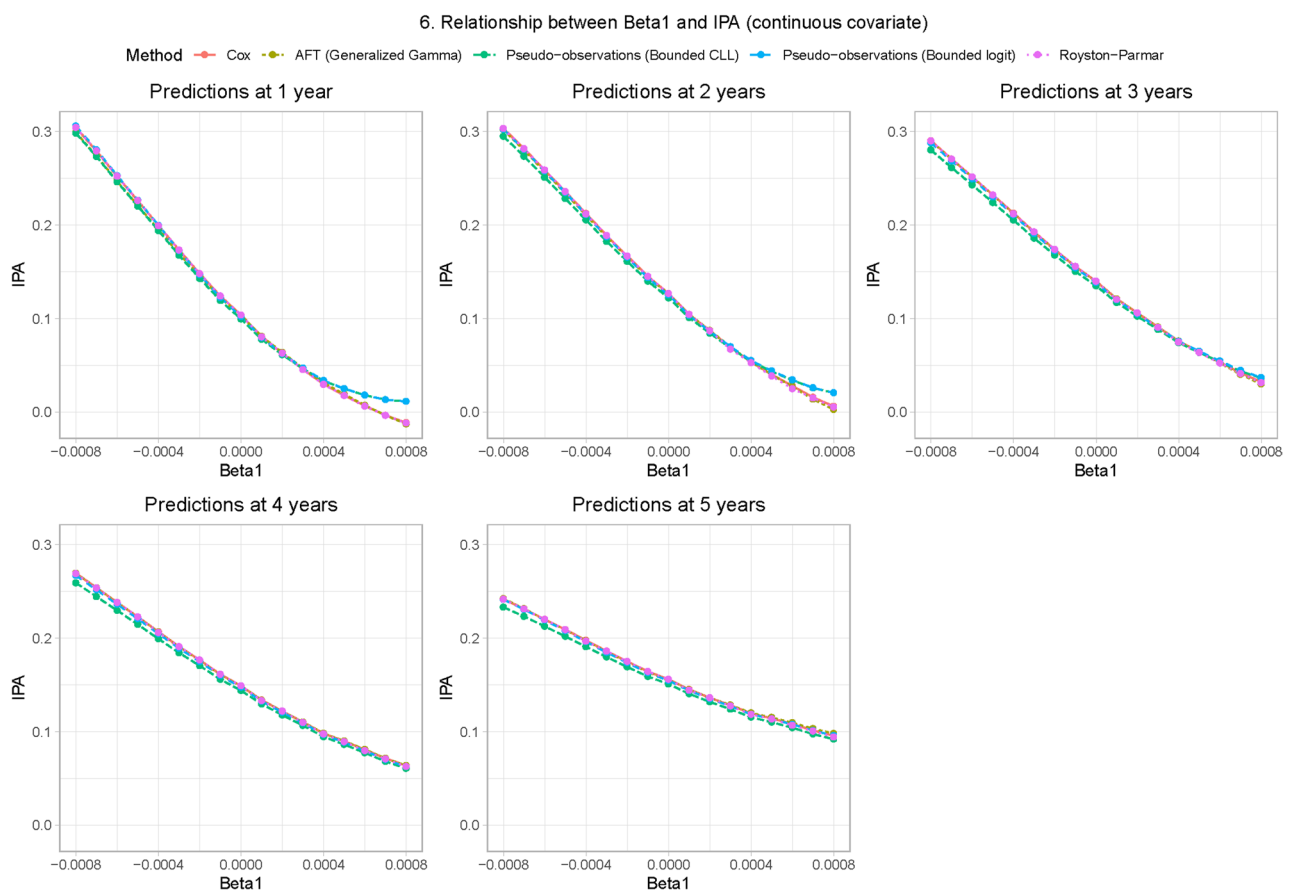


Fig. 6 Relationship between Beta1 and IPA (continuous covariate)

The computed IPAs from all five models displayed similar variability across simulation replicates.

Sensitivity analysis – the effect of prevalence of the binary covariate

In the above simulations, the binary covariate was simulated to have a prevalence of 0.5 ($\Pr(X_2=1)=0.5$). In this section we conducted a sensitivity analysis in which we assessed the impact of varying the prevalence of the binary variable for which the proportional hazards assumption was violated. In these simulations we made the following two modifications to the simulations described in Sect. 2: (i) we fixed $\beta_1 = 0.0004$, thereby inducing a moderate violation of the proportional hazards assumption; (ii) we allowed $\Pr(X_2=1)$ to vary from 0.05 to 0.50 in increments of 0.05. We only considered settings in which the proportional hazards assumption was violated for the binary variable. The statistical analyses in the simulated datasets were identical to those described in Sect. 2.

The relation between the prevalence of the binary variable ($\Pr(X_2=1)$) and Harrell’s concordance statistic is reported in Fig. 7. The prevalence of the binary variable for which the proportional hazards assumption was

violated had essentially no impact on overall discrimination (note the scale of the vertical axis).

The relation between the prevalence of the binary variable ($\Pr(X_2=1)$) and time-varying AUC is reported in Fig. 8. The prevalence of binary variable had at most a very modest impact on discrimination when the prediction horizon was short and at most a minor impact at longer prediction horizons.

The relation between the prevalence of the binary variable ($\Pr(X_2=1)$) and the time-varying IPA is reported in Fig. 9. The prevalence of binary variable had at most a modest impact on time-varying IPA, with the magnitude of the effect of the prevalence of the binary variable increasing modestly at the length of the prediction horizon increased.

Discussion

We examined the impact of violation of the proportional hazards assumption on the discrimination of a fitted Cox regression model in settings in which the time-varying log-hazard ratio was a linear function of time. We observed an inverse relationship between the regression coefficient relating the time-varying log-hazard ratio with time and concordance (as assessed

7. Relation between prevalence of binary variable and concordance

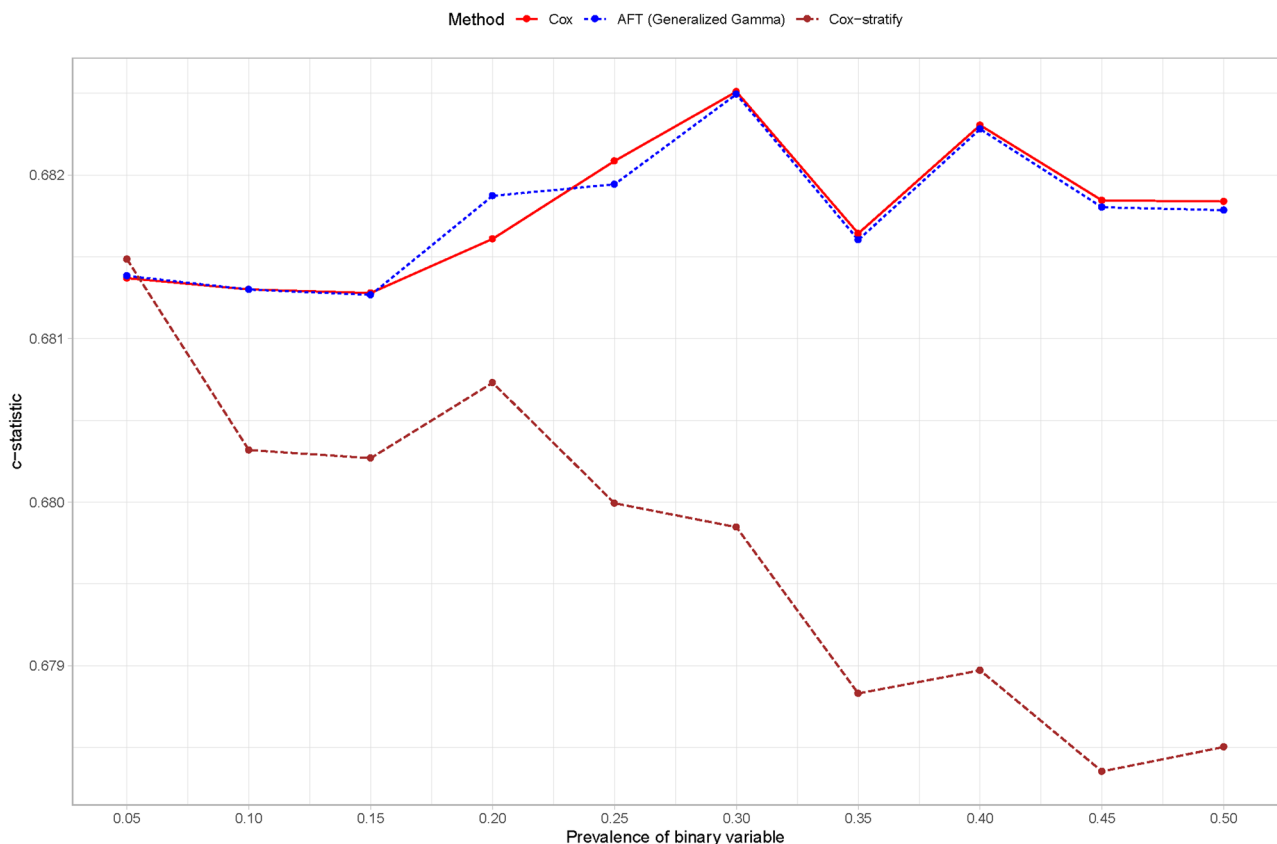


Fig. 7 Relation between prevalence of binary variable and concordance

using Harrell’s concordance statistic), discrimination (as assessed using the time-varying AUC), and predictive accuracy (as assessed using the time-varying IPA). Thus, concordance, discrimination, and predictive accuracy were highest when the magnitude of the violation of the proportional hazards assumption was very strong and the time-varying hazard ratio was initially very large and then decreased over time. Conversely, concordance, discrimination, and predictive accuracy were lowest when the magnitude of the violation of the proportional hazards assumption was very strong and the time-varying hazard ratio was initially low and then increased over time. Similar findings were observed for alternative prediction methods: a parametric AFT survival model with a generalized gamma distribution, Royston and Parmar’s spline-based parametric survival models, and generalized linear models using pseudo-observations.

There is a paucity of research on the impact of the violation of the proportional hazards assumption on the discrimination of predictions obtained from a Cox regression model. Royston and Altman, when discussing principles for the external validation of a clinical prediction model developed using Cox regression, suggested that, even if the proportional hazards assumption is

violated, a Cox model ‘may still provide good discrimination’ [6]. Our findings suggest this is not necessarily the case. Depending on the prediction horizon and on whether the proportional hazards assumption was violated for a binary variable or for a continuous variable, violation of the proportional hazards assumption can result in a substantial change in model performance. For example, when the proportional hazards assumption was violated for a continuous variable and with a one-year prediction horizon, the AUC was approximately 0.55 when $\beta_1 = 0.0008$, whereas the AUC was approximately 0.70 when the proportional hazards assumption was satisfied. While the latter could be described as denoting moderate discrimination, the former would be described as denoting very poor discrimination. However, if the violation of the proportional hazards assumption is mild, then a fitted Cox model that ignores this violation will tend to have discrimination comparable to that which would be observed had the proportional hazards assumption been satisfied.

The primary limitation of the current study is its reliance on Monte Carlo simulations. The simulations were computationally intensive because of the complexity of the data-generating process. We restricted the number of

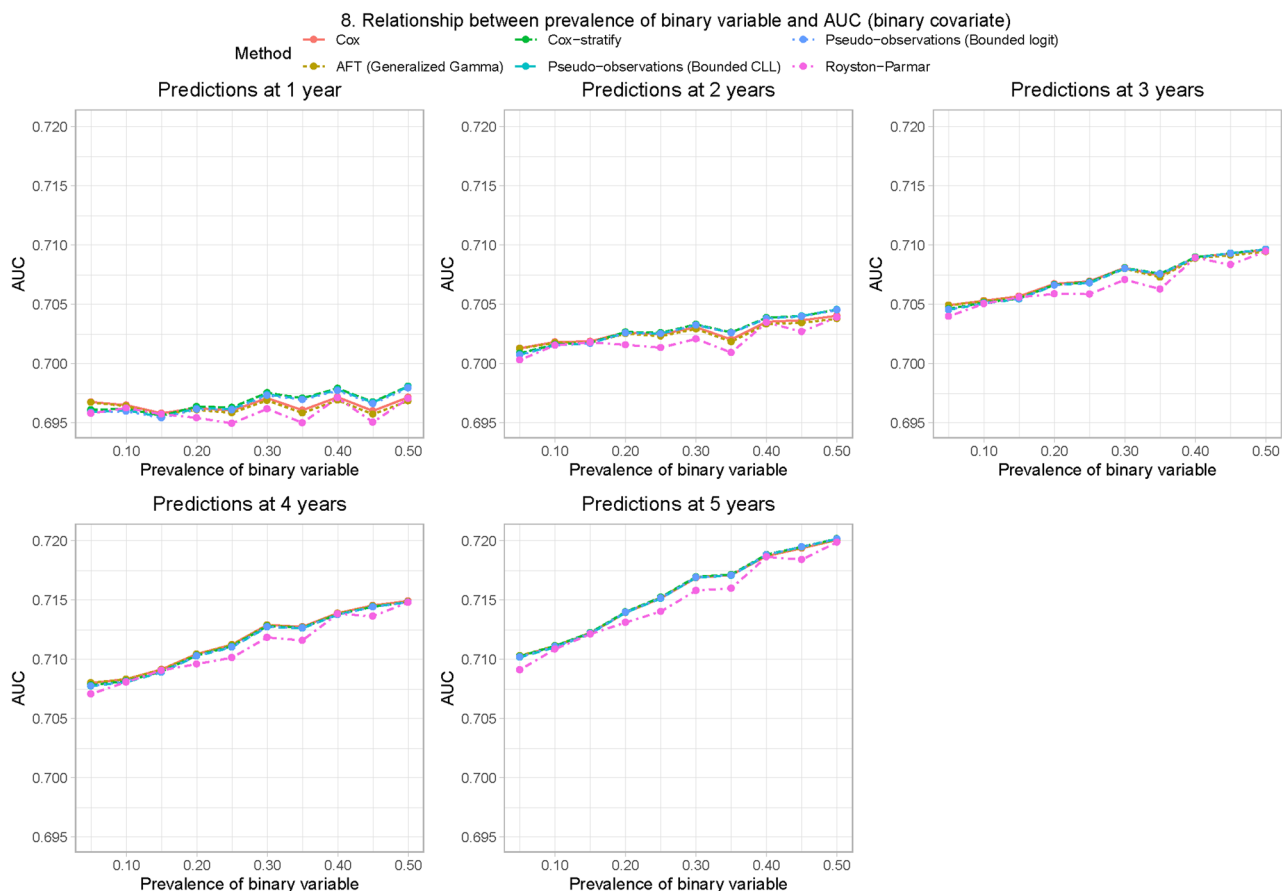


Fig. 8 Relationship between prevalence of binary variable and AUC (binary covariate)

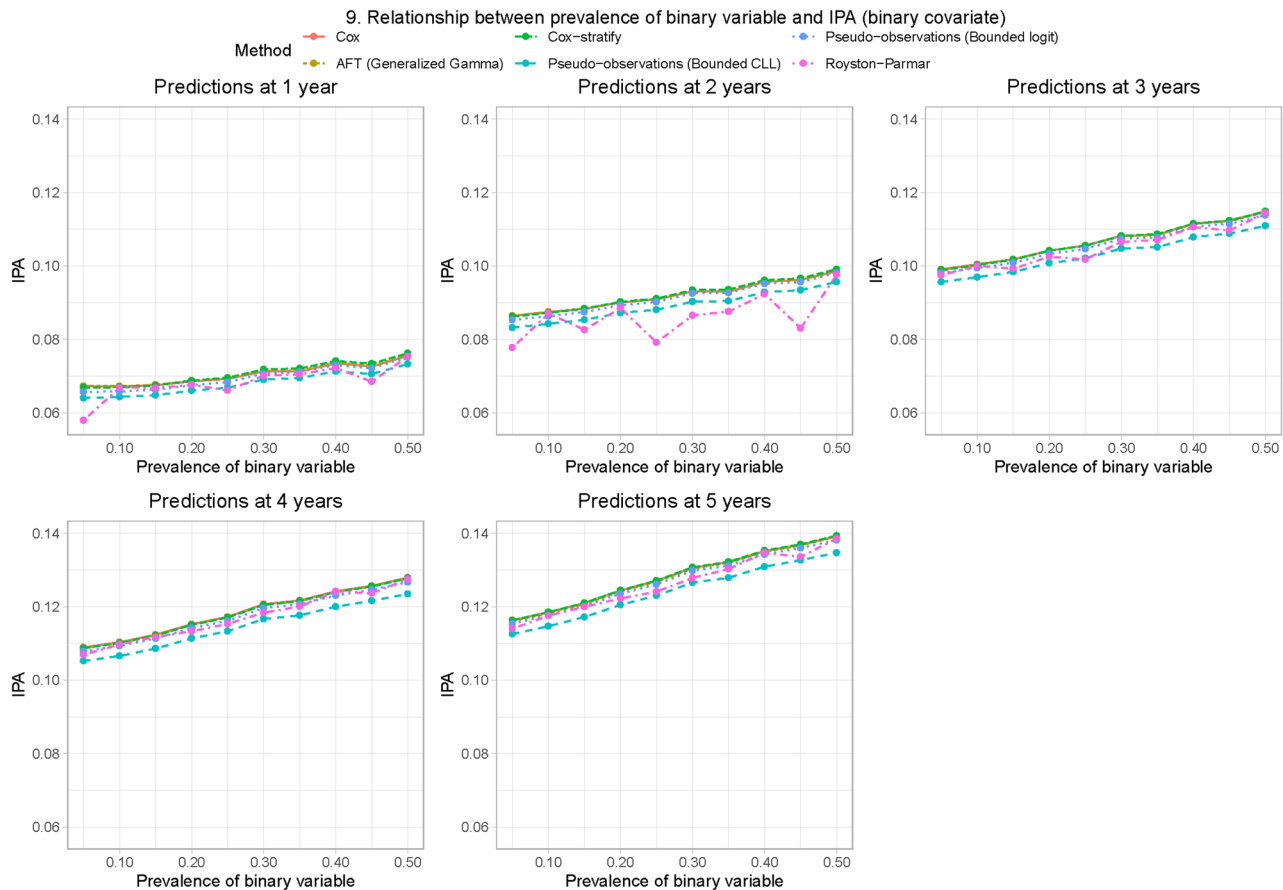


Fig. 9 Relationship between prevalence of binary variable and IPA (binary covariate)

scenarios that we examined because the use of a Weibull model to simulate time-to-event outcomes is computationally intensive as it requires using numerical integration of the cumulative hazard function and root-finding techniques to invert the cumulative hazard function. Despite this, we examined 34 different scenarios, characterized by the magnitude of the violation of the proportional hazards assumption and by the type of variable for which the proportional hazards assumption was violated. A second limitation was that in our simulations, we simulated outcomes using a Weibull parametric survival model. It is possible that different results would be observed when different models were used for generating outcomes. A third limitation is that our study did not include an empirical analysis but relied solely on Monte Carlo simulations. While we could have used the different methods in an empirical analysis of data in which it was known that the proportional hazards assumption was violated, this would only have allowed us to compare the discrimination of the different methods. It would not have allowed us to compare the observed discrimination with the discrimination that would have been observed had the proportional hazards assumption not been violated. This latter comparison was the focus of the study:

the impact of the violation (and the magnitude of the violation) of the proportional hazards assumption on the discrimination of predictions obtained from the Cox model. Such a question cannot be addressed using empirical analyses.

Conclusion

Violation of the proportional hazards assumption can have a meaningful impact on the discrimination of a Cox regression model. Compared to settings in which the proportional hazards assumption was satisfied, discrimination and predictive accuracy decreased in settings in which the log-hazard ratio was positively associated with time. Conversely, compared to settings in which the proportional hazards assumption was satisfied, discrimination and predictive accuracy increased in settings in which the log-hazard ratio was negatively associated with time. Compared with the use of a Cox regression model, the use of AFT parametric survival models, Royston and Parmar's spline-based parametric survival models, and generalized linear models using pseudo-observation did not result in estimates with improved discrimination in settings in which the proportional hazards assumption was violated.

Abbreviations

AFT	Accelerated failure time
AUC	Area under the curve
IPA	Index of prediction accuracy
ROC	Receiver operating characteristic

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41512-026-00223-0>.

Supplementary Material 1.

Acknowledgements

Not applicable.

Authors' contributions

PCA conceived the study, design and conducted the simulations and statistical analyses, and drafted the manuscript. DG contributed to study design and revised the manuscript for important intellectual content. All authors approved the final manuscript.

Funding

ICES is an independent, non-profit research institute funded by an annual grant from the Ontario Ministry of Health (MOH) and the Ministry of Long-Term Care (MLTC). This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health (MOH) and the Ministry of Long-Term Care (MLTC). This study also received funding from the Canadian Institutes of Health Research (CIHR) (PJT 183898). Daniele Giardiello is funded by the National Plan for NRRP Complementary Investments (PNC, established with the decree-law 6 May 2021, *n. 59, converted by law n. 101 of 2021*) in the call for the funding of research initiatives for technologies and innovative trajectories in the health and care sectors (Directorial Decree n. 931 of 06-06-2022) - project n. PNC0000003 - Advancing Technologies for Human-centred Medicine (project acronym: ANTHEM). The analyses, conclusions, opinions and statements expressed herein are solely those of the authors and do not reflect those of the funding or data sources; no endorsement is intended or should be inferred.

Data availability

Only synthetic data generated using data-generating processes were used in this study. The data-generating processes are described in the paper and can be reproduced by the reader.

Declarations

Ethics approval and consent to participate

No data were used in generating the results for this study. The study employed Monte Carlo simulations in which synthetic data were generated using data-generating processes. As such, no ethics approval or consent to participate were required.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 31 July 2025 / Accepted: 5 February 2026

Published online: 12 February 2026

References

1. Aalen OO, Borgan O, Gjessing HK. Survival and event history analysis. New York, NY: Springer; 2008.
2. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–38.
3. Steyerberg EW. Clinical prediction models. Second ed. New York: Springer; 2019.
4. Harrell FE Jr. Regression modeling strategies. Second ed. New York, NY: Springer; 2015.
5. Austin PC, Giardiello D. The impact of violation of the proportional hazards assumption on the calibration of the Cox proportional hazards model. *Stat Med*. 2025;44(13–14):e70161.
6. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol*. 2013;13:33.
7. Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. *Stat Med*. 2013;32(23):4118–34.
8. Harrell FE Jr, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med*. 1984;3(2):143–52.
9. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics Med*. 2011;30(10):1105–17.
10. Blanche P, Kattan MW, Gerds TA. The c-index is not proper for the evaluation of t-year predicted risks. *Biostatistics*. 2019;20(2):347–57.
11. Kattan MW, Gerds TA. The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. *Diagn Progn Res*. 2018;2:7.
12. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and Estimation of treatment effects. *Stat Med*. 2002;21(15):2175–97.
13. Andersen PK, Perme MP. Pseudo-observations in survival analysis. *Stat Methods Med Res*. 2010;19(1):71–99.
14. Therneau T. Pseudo-values for survival data. 2024; <https://cran.r-project.org/web/packages/survivalVignettes/vignettes/pseudo.html>. Accessed November 6, 2024, 2024.
15. Team RCD. *R: a Language and environment for statistical computing* [computer program]. Vienna: R Foundation for Statistical Computing. 2005.
16. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38(11):2074–102.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.