

Why Generating No Evidence is Not a Good Solution When Encountering Zero Events: Reducing Bias From Monotone Likelihood in Rare-Outcome Cox Models Using Penalization

Guy Cafri¹
Martijn Schuemie¹

1. Global Epidemiology Organization, Johnson and Johnson, Titusville, New Jersey, USA

Background

Estimating comparative safety in pharmacoepidemiology often involves time-to-event analyses with rare outcomes. In post-marketing surveillance settings where comparative study designs are used, it is common to encounter analyses with very few - or even zero - observed events in one treatment group. Standard partial-likelihood estimation for Cox regression can behave poorly in these settings. Most problematic is a situation in which the likelihood keeps increasing as the parameter estimate approaches $\pm\infty$, also known as monotone likelihood. We suspect that a common practice in this situation is simply not to report the results of an analysis where this problem arises because a conventional Cox model cannot produce an informative result. This seemingly innocent practice can lead to considerable downward bias in the presence of a true safety effect, wherein the target exposure has increased risk. Although methods for combining Cox regression results exist for multi data-source studies when there are small or zero counts [1], methods for inference from a single data source with very few or zero events are needed.

In this study, Monte Carlo simulations are used to illustrate: 1) the magnitude of bias in the treatment effect estimate when the practice consists of not reporting results from a conventional Cox model when a monotone likelihood exists, and 2) the magnitude of bias in the treatment effect estimate when a penalized solution [2,3] is applied to settings when a monotone likelihood is encountered.

Methods

Three data generation factors were manipulated: outcome incidence (0.01%, 0.1%, 0.5% or 1%) the hazard ratio (1, 2, 4 or 8), and the sample size (1,000, 2,500, 5,000 or 10,000). The outcome for the study is percent relative bias.

Data generation for each condition was based on generating population-level data of 10,000,000 survival times, with 5,000,000 in each of the treatment and comparison groups. From this large population-level data generated for each simulated condition, 1,000 simulated datasets were drawn with each having sample sizes of 1,000, 2,500, 5,000 or 10,000. For individuals in each group we generated an event time and a censoring time. Censoring times were generated in the same way for both treatment and comparison groups, but the event time generation varied between the groups. We generated censoring and event times using the approach described by Bender and colleagues [4].

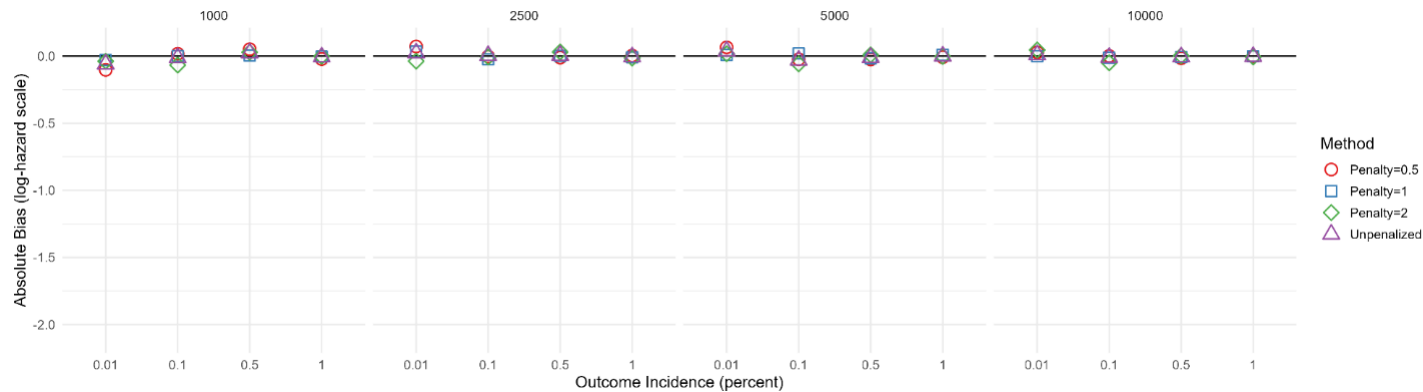
Four methods of point estimation were evaluated. For all four methods, regression models included the treatment indicator as the only predictor. The regression coefficient from the model, $\hat{\beta}$, is the point

estimate used to evaluate bias. The first method of point estimation was based on a standard Cox regression model. In the presence of a monotone partial likelihood, the point estimate was set to missing. The remaining three methods implemented Firth's correction in the presence of data resulting in a monotone partial likelihood, with penalty (λ) of 0.5, 1.0 or 2.0 (higher penalties will lead to smaller effects). When the data did not result in a monotone partial likelihood the standard Cox regression model was retained. Percent relative bias is calculated as: $100 * (\bar{\beta} - \beta) / \beta$, where β is $\ln(HR)$ and $\bar{\beta} = \sum_{s=1}^S \hat{\beta}_s / S$ ($s = 1, \dots, S$; $S = 1000$ simulated data sets).

Results

Across scenarios with a null effect ($HR = 1$), absolute bias on the log-hazard scale was uniformly small (Figure 1). The hybrid estimator—using the standard Cox model when the partial likelihood was well-behaved and switching to the Firth-penalized Cox when monotone likelihood arose—performed comparably to the unpenalized estimator. Differences among $\lambda = 0.5, 1$, and 2 were small.

Figure 1. Bias by Sample Size and Outcome Incidence with Hazard Ratio of 1



When the true effect was moderate ($HR = 2$), percent relative bias displayed a clear dependence on the method of estimation, sample size and outcome incidence. The practice of fitting a standard Cox model and effectively discarding results with monotone likelihood produced substantial downward (attenuation) bias in the rarest-outcome and smallest-sample scenarios, with bias magnitude decreasing as outcome incidence and sample size increased. In contrast, the hybrid estimator markedly reduced attenuation in scenarios with small sample size and/or low outcome incidence, with $\lambda = 0.5$ generally yielding the smallest bias and the largest penalty of $\lambda = 2$ tending to over-shrink the effect in the same conditions but still improving upon the unpenalized approach. Notably, when the outcome incidence was at its lowest (0.01%), penalization yielded no improvement when the sample size was 1,000 or 2,500, and despite modest improvements when sample size was 5,000 or 10,000, the degree of relative bias was still large. At the other extreme, when the outcome incidence was at its highest (1%), all methods were comparable and had little bias. When the outcome incidence was in between these extremes, the degree to which the hybrid method was able to reduce bias relative to the standard Cox model depended on the sample size, with notable improvements with the lowest, but relatively little improvements with larger sample sizes. For larger effects the patterns described above were more pronounced (Figures 3).

Figure 2. Bias by Sample Size and Outcome Incidence with Hazard Ratio of 2

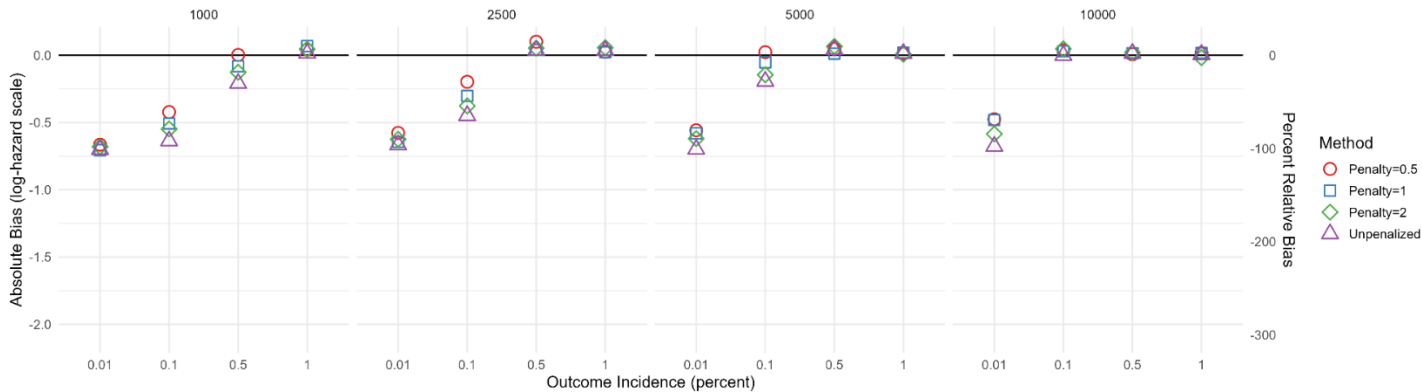
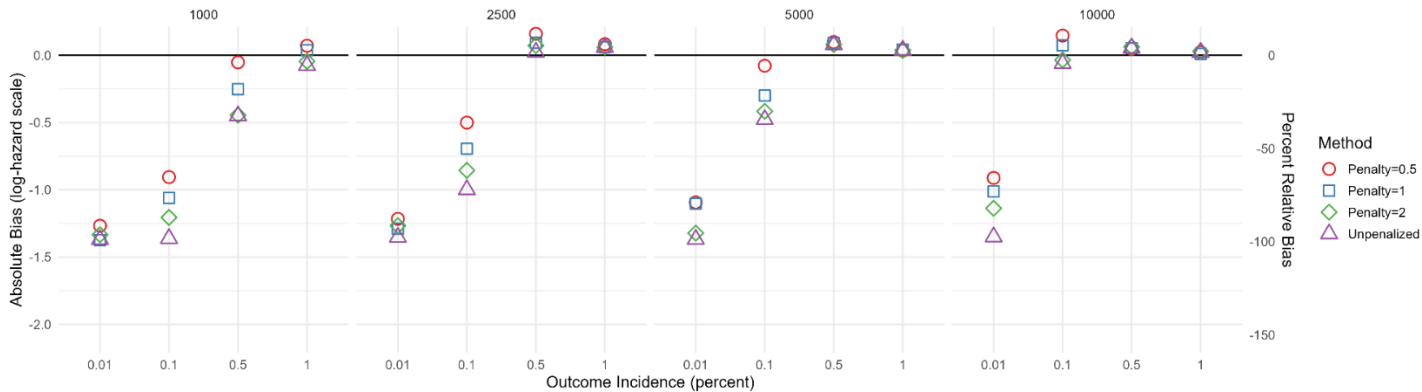


Figure 3. Bias by Sample Size and Outcome Incidence with Hazard Ratio of 4



Conclusion

Our simulations show that in rare-outcome, small-sample studies, fitting a conventional Cox model and omitting analyses with monotone likelihood produces substantial downward bias, which intensifies as effects grow and events become rarer. A hybrid strategy—using the standard Cox model when the partial likelihood is well-behaved and switching to a Firth-penalized Cox when monotone likelihood arises—markedly reduces this attenuation. In the presence of true treatment effects, lighter penalization generally yielded the smallest bias, while stronger penalization modestly over-shrank but still improved upon the unpenalized approach; however, when incidence was extremely low (0.01%) and samples were small (1,000–2,500), bias remained large regardless of the penalty. We recommend adopting the hybrid approach evaluated in this study to avoid the bias inherent in not reporting results when monotone likelihood is encountered, while recognizing that extremely rare outcomes may require larger samples or alternative designs.

References

- 1.Schuemie, M.J., Chen, Y., Madigan, D., & Suchard, M.A. (2021). Combining Cox Regressions Across a Heterogeneous Distributed Research Network Facing Small and Zero Counts. *Statistical Methods in Medical Research*, 31(3), 438–450
- 2.Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993;80:27-38.
- 3.Heinze G, Schemper M. A solution to the problem of monotone likelihood in Cox regression. *Biometrics*. 2001;57(1):114-119.
- 4.Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine* 2005; 24:1713–1723.