Review Report for “Identifying Critical Shortcomings of Estimators of Discriminative Performance in Time-to-Event Analyses: A Comparison Study”

The paper investigates potential issues with estimators of discrimination used to evaluate the performance of Cox models. Specifically, the authors argue that existing estimators can lead to overestimation of out-of-sample discriminative performance in certain situations, even when the model is correctly specified. The paper also provides alternative estimators and illustrate the behavior of different estimators through a simulation study and a data application.

Major comments

1. The estimator in Equation (2) may not be applicable when the survival time has a continuous distribution, since the probability of $T = t$ for a given t is negligible. Consequently, it may not be possible to establish large sample properties such as consistency. In light of these limitations, it may not be meaningful to consider this estimator in the context of time-to-event models with continuous survival times.

2. The authors note that the Heagerty-Zheng estimator may not perform well in situations where the model has a complex form with a large number of free parameters. However, this may be due to the fact that the estimator for the Cox model assumes a fixed number of covariates and may exhibit poor performance in high-dimensional predictor settings. In such cases, it may be more appropriate to estimate the model coefficients using penalized Cox regression techniques such as lasso or elastic net, rather than using the original Cox model. By incorporating penalties on the size of the coefficients, these methods can effectively handle high-dimensional settings and may improve the performance of the Heagerty-Zheng estimator.

3. Can the authors provide more details about the smoothed non-parametric estimators using penalized regression splines? I found the equation on page 4, lines 50-52, to be difficult to understand, and some of the notations, such as $\varepsilon(t)$ and $B_k(t)$, seem to be undefined. Furthermore, while the left-hand side of the equation appears to be an estimator, it is unclear how the right-hand side depends on the data.

   Similarly, I also found the notation used in the equation on page 5, lines 3-5, to be unclear. The repeated use of $\varepsilon(t)$ and $B_k(t)$ makes it difficult to determine if they are the same as in the previous equation.

4. It would be beneficial for the authors to provide some theoretical justification for the proposed estimators, as this would add to the overall strength and validity of their

findings. While the simulation studies presented in the paper provide insights into the potential limitations of certain estimators, they may not be sufficient to demonstrate the good performance of the proposed estimators in general applications. A more rigorous theoretical analysis of the proposed estimators could provide additional support for their reliability and improve the overall credibility of the paper's conclusions.

5. While the paper notes that the Heagerty-Zheng estimator are prone to overestimate the model performance when the covariate space of new samples has little overlap with the training sample, it does not appear that the authors presented simulation studies to demonstrate this point (if I did not overlook anything).

   Moreover, the term "little overlap" in covariate spaces may be somewhat vague. It would be helpful if the authors could provide more specific information on what they mean by this term and why it leads to overestimation of model performance.

Minor comments

1. Page 4, line 46-47: $2\hat{f}(t)\hat{S}(t)/1 - \hat{S}^2(\tau)$ should be $2\hat{f}(t)\hat{S}(t)/\{1 - \hat{S}^2(\tau)\}$?
2. The notation η_i seems confusing. It is unclear whether it represents estimated value of $X_i^t\beta$ or $X_i^t\beta$ itself.