

EVALUATING THE ACCURACY OF A PREDICTIVE MODEL IN SURVIVAL ANALYSIS

Chaya S. Moskowitz and Michael W. Kattan
Memorial Sloan-Kettering Cancer Center, New York NY

Background. Risk prediction models can be used as an aid when determining patient management. Because incorrect predictions could have substantial consequences, it is very important to evaluate the predictive accuracy of a model. When the risk prediction model is the result of a survival analysis where the outcomes are potentially censored, there is no standard way of evaluating predictive accuracy. There have been many different measures proposed for this purpose, but they all have limitations. The most frequently used measure in practical applications is the concordance index (c-index) proposed by Harrell et al. (1982). The c-index, however, is biased when the data are censored, a problem that does not appear to be widely recognized. The purpose of this poster is to highlight and explore this bias.

Methods. To study the properties of the c-index when data are censored, we conducted a simulation study. We simulated failure times and predictions from a risk model by generating data from a binormal distribution, which allowed us to prespecify the true level of concordance. We generated censoring times from a uniform distribution. By varying the various distribution parameters, the true level of concordance, the percentage of censored data, and the sample size, we were able to explore how well the c-index estimates true concordance in a wide variety of situations.

Results. We found evidence of bias in the estimate of the c-index across all sample sizes studied. The degree of bias increases as the percentage of data that is censored increases and appears to be the largest for moderate levels of (true) concordance. In the situations we explored, the average bias across simulations ranged from close to zero up to 0.07 when 75% of the data was censored. The direction of the bias is usually positive, indicating that, in general, the c-index overestimates true concordance, thereby estimating the predictive accuracy of the risk model to be higher than it is in truth.

Conclusions. While the c-index is often used to evaluate the accuracy of risk prediction models, when there is a high degree of censoring, it is a limited measure of predictive accuracy. Further work needs to be done to develop a measure of predictive accuracy that is intuitively interpretable and not affected by censoring.

References

Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982;**247**(18):2543-2546.

Schemper M, Stare J. Explained variation in survival analysis. *Statistics in Medicine* 1996;**15**:1999-2012.