Optimal Data Analysis
Vol. 9 (January 13, 2020), 3-6
Copyright 2020 by Optimal Data Analysis, LLC
2155-0182/10/$3.00

Selecting an Appropriate Weighting Strategy in Maximum-Accuracy Time-to-Event (Survival) Analysis

Paul R. Yarnold, Ph.D., Nathaniel J. Rhodes, Pharm.D., M.Sc.,
Optimal Data Analysis, LLC
Chicago College of Pharmacy, and the
Pharmacometrics Center of Excellence, Midwestern University

and

Ariel Linden, Dr.P.H.
Linden Consulting Group, LLC

Different weighting schemes in optimal survival analysis are considered.

Consider an optimal[1,2] survival analysis[3,4] for a sample of subjects observed for a finite number of consecutive time periods: 1, 2, 3 … X.

If observations are weighted by X_i then the non-censored surviving observations which avoided the outcome for X periods are weighted more heavily than observations that achieved the outcome in fewer than X periods, all other things being equal (e.g., this does *not* assume a complex weight, such as the value X_i multiplied by a propensity score[5]). That is, in this scheme observations that *did NOT achieve* the outcome are weighted most heavily. For example, if the outcome is death, then observations that lived the entire X periods are weighted more heavily than observations that perished in fewer than X periods. This weighting scheme favors accurate classification of *surviving* observations.

In contrast, if observations are weighted by $1/X_i$ then non-censored observations that did not survive (i.e., achieved the outcome in fewer than X periods) are weighted more heavily than observations that failed to achieve the outcome in X periods, all other things being equal. That is, in this scheme observations that *did achieve* the outcome are weighted the most heavily. For example, if the outcome is death, then observations that perished in fewer than X periods are weighted more heavily than observations that perished (or failed to perish) in X periods. This weighting scheme favors accurate classification of *non-surviving* observations.

Finally, consider analysis that observes a sample of subjects over a fixed time period of *exactly* X time periods, which does *not* weight observations by time-to-event, but rather instead weights all observations equally by unity (or by prior odds[1,2], by a propensity score[6,7], and/or by a measure of value[8-11]). The objective of such an analysis is to contrast observations who live *vs*.

Copyright 2020 by Optimal Data Analysis, LLC
2155-0182/10/$3.00

perish over X or fewer units of time. This weighting scheme does not favor the accurate classification of *surviving vs*. *non-surviving* observations, but instead provides a means of ensuring that the treatment groups are comparable.[12-14]

Most classification models seek balanced sensitivity and specificity so as to maximize ROC area.[15-20] As always, the crucial question is—is the model fit for purpose? Thus, the decision to weight outcomes more heavily *vs*. non-outcomes, to do the opposite, or to weight observations equally is entirely a question of whether the goal is to predict the outcomes or the non-outcomes most accurately. A related caveat is the problem familiar to data scientists—the rare event case caused by base rates of <10% of the sample: in this scenario the event rate is low and often the "best guess" for an empirical model is to guess "non-event".[21] There are many alternative ways that this problem may be addressed—ranging from complex simulation based approaches (e.g., SMOTE) to the random under- and over-sampling of training datasets.[22] However, we believe that simpler weighting schemes (e.g., at risk time, propensity, value) are highly intuitive and thus offer translational clarity for end-users. That is, rather than presenting "black box" solutions for class imbalance, the model offers insight into the factors that end-users should consider as being more influential over the final outcome classification. Finally, it is important to consider the "bias" of predictions: is a model with significant bias acceptable for use? To answer this, an end-user must consider the consequences of false positives and false negatives.[23] The impact of false signals is clearly relevant to the clinical or use case scenario. For example, consequences of a false positive prediction for invasive breast cancers based on mammogram, antibiotic resistance in pneumonia based on prior hospitalization history, bridge stability according to electrical conductance from wire supports, or hostile inbound aircraft using radar signals are all very different from a false positive prediction for a tooth cavity based on dental X-ray imaging data.

Weighting strategies discussed presently do different things, for different purposes. The propensity score and similar types of weights, for example prior odds, are used to ensure that treatment groups (or exposure groups in epidemiology) are comparable and are therefore exchangeable. Weighting by value is used to ensure that differences between observations are accounted for in the hopes of maximizing the omnibus positive gain (return) or minimizing the omnibus negative gain (loss) for a group. And, weighting by time is used to understand the etiology of an outcome: for example, in some cases the longer subjects are monitored the more likely they are to contract an illness or die, whereas in other cases subjects may be more likely to contract the illness or die relatively quickly—and thus the longer that they are monitored, the more likely it is that they have passed through the period of risk. These types of adjustments may be made using parametric survival models whereby a researcher chooses the distribution model believed to best represent the trajectory of acquiring the outcome. In contrast, maximum-accuracy methods require no distributional assumptions, and instead simply identify the model(s) which explicitly maximize model accuracy in predicting the outcome being investigated.

An empirical demonstration of these alternative weighting schemes is welcomed.

References

[1]Yarnold PR (2017). What is optimal data analysis? *Optimal Data Analysis, 6*, 26-42.

[2]Yarnold PR, Soltysik RC (2016). *Maximizing Predictive Accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286

Copyright 2020 by Optimal Data Analysis, LLC
2155-0182/10/$3.00

[3]Linden A, Yarnold PR (2017). Modeling time-to-event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice*, *23*, 1299-1308. DOI: 10.1111/jep.12779

[4]Linden A, Yarnold PR (2018). Using ODA in the evaluation of randomized controlled studies: Application to survival outcomes. *Optimal Data Analysis*, *7*, 50-53.

[5]Linden A, Yarnold PR (2018). Estimating causal effects for survival (time-to-event) outcomes by combining classification tree analysis and propensity score weighting. *Journal of Evaluation in Clinical Practice*, *24*, 380-387. DOI: 10.1111/jep.12859

[6]Linden A, Yarnold PR (2017). Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice*, *23*, 703-712. DOI: 10.1111/jep.12744

[7]Yarnold PR, Linden A (2017). Computing propensity score weights for CTA models involving perfectly predicted endpoints. *Optimal Data Analysis*, *6*, 43-46.

[8]Yarnold PR (2019). Maximum-precision Markov transition table: Successive daily change in closing price of a utility stock. *Optimal Data Analysis*, *8*, 3-10.

[9]Yarnold PR, Soltysik RC (2019). Confirming the efficacy of weighting in optimal Markov analysis: Modeling serial symptom ratings. *Optimal Data Analysis*, *8*, 53-55.

[10]Yarnold PR (2019). Weighted optimal Markov model of a single outcome: Ipsative standardization of ordinal ratings is unnecessary. *Optimal Data Analysis*, *8*, 60.

[11]Yarnold PR (2019). Optimal Markov model relating *two* time-lagged outcomes. *Optimal Data Analysis*, *8*, 61-63.

[12]Linden A, Yarnold PR (2016). Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice*, *22*, 848-854.

[13]Linden A, Yarnold PR (2016). Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, *22*, 868-874.

[14]Linden A, Yarnold PR (2016). Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments. *Journal of Evaluation in Clinical Practice*, *22*, 875-885.

[15]Rhodes NJ, Kuti JL, Nicolau DP, Van Wart S, Nicasio AM, Liu J, Lee BJ, Neely MN, Scheetz MH (2015). Defining clinical exposures of cefepime for gram-negative bloodstream infections that are associated with improved survival. *Antimicrobial Agents and Chemotherapy*, *60*, 1401-1410.

[16]DiPippo AJ, Tverdek FP, Tarrand JJ, Munita JM, Tran TT, Arias CA, Shelburne SA, Aitken SL (2017). Daptomycin non-susceptible *Enterococcus faecium* in leukemia patients: Role of prior daptomycin exposure. *The Journal of infection*, *74*(3), 243–247. doi:10.1016/j.jinf.2016.11.004

[17]Yarnold PR, Hart LA, Soltysik RC (1994). Optimizing the classification performance of logistic regression and Fisher's discriminant analyses. *Educational and Psychological Measurement*, *54*, 73-85.

[18]Yarnold PR, Linden A (2019). Optimizing suboptimal classification trees: Matlab® CART model predicting probability of lower limb prosthesis user's functional potential. *Optimal Data Analysis*, *8*, 84-93.

Copyright 2020 by Optimal Data Analysis, LLC
2155-0182/10/$3.00

[19]Yarnold PR (2019). Optimizing suboptimal classification trees: S-PLUS® propensity score model for adjusted comparison of hospitalized *vs*. ambulatory patients with community-acquired pneumonia. *Optimal Data Analysis*, *8*, 38-47.

[20]Yarnold PR (2019). More on: "Optimizing suboptimal classification trees: S-PLUS® propensity score model for adjusted comparison of hospitalized *vs*. ambulatory patients with community-acquired pneumonia". *Optimal Data Analysis*, *8*, 56-59.

[21]Prati RC, Batista GEAPA, Silva DF (2015). Class imbalance revisited: A new experimental setup to assess the performance of treatment methods. *Knowledge and Information Systems*, *45*, 247-270.

[22]Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321-357. DOI:https://doi.org/10.1613/jair.953

[23]Yarnold PR (2014). UniODA *vs*. ROC analysis: Computing the "optimal" cut-point. *Optimal Data Analysis*, *3*, 117-120.

Author Notes

No conflict of interest was reported.