Advances in Methodology and Statistics / Metodološki zvezki, Vol. 18, No. 1, 2021, 31–37
https://doi.org/10.51936/atez3037

A note on generalisations of the concordance index for survival data

Nataša Kejžar*, Janez Stare

University of Ljubljana, Faculty of Medicine, Ljubljana, Slovenia

Abstract

Concordance index (c-index) was adapted to survival data by Harrell (1982). In its basic form, the index depends on censoring, however the issue can be effectively dealt with. More importantly, Harrell's c-index cannot be used with time-varying effects and/or time-dependent covariates, and several generalisations were proposed. We look at some of them, explore their differences, point to a basic difference between these generalisations, and strongly favour one type of generalisation.

Keywords: survival analysis, models with time-dependent covariates/time-varying effects, c-index

1. Introduction

Concordance parameter defines the probability of selecting a concordant pair from the population. Without considering time-to-event data this parameter is also known under the name of probability index (Acion et al., 2006), and it is identical to the area under the curve (AUC) measure (Hanley & McNeil, 1982). Its estimator (see Lehmann, 1951) is consistent and unbiased with minimum variance; and scaled version of this estimator is used as a test statistic for the non-parametric Mann-Whitney U test (Lehmann, 1951).

One of commonly used estimators for the concordance parameter for survival data is the c-index. There have been some suggestions (Antolini et al., 2005; Gerds et al., 2013; Kremers, 2007) on how to generalise the c-index since Harrell et al. (1982) first adapted it for use for event history data. The concordance estimation parameter can be also obtained using the estimate for the area under the ROC curve for which multiple generalisations have been proposed (Chambless & Diao, 2006; Heagerty & Zheng, 2005). The common goal of such generalisations is to adapt the measure to time-varying effects and time-dependent covariates. None of these suggestions (Antolini et al., 2005; Chambless & Diao, 2006; Gerds et al., 2013; Heagerty & Zheng, 2005; Kremers, 2007) defines what is meant by generalisation, possibly

*Corresponding author

Email addresses: natasa.kejzar@mf.uni-lj.si (Nataša Kejžar), janez.stare@mf.uni-lj.si (Janez Stare)

ORCID iDs: 0000-0003-3069-9863 (Nataša Kejžar), 0000-0002-2564-8781 (Janez Stare)

because the idea is so obvious that it doesn't have to be explicitly formulated. Still, a clear definition helps to distinguish generalisations from modifications. In this brief note we first define what we believe should be understood as a generalisation of the c-index, and then discuss some proposals in light of this definition, compare them, and strongly argue for one type of generalisations.

C-index for survival data (Harrell et al., 1982) is defined as follows

$$c = \frac{\text{\# concordant pairs}}{\text{\# comparable pairs}}. \tag{1.1}$$

A pair is concordant if predicted survival times for the pair are in the same order as observed survival times.

If there is no censoring in the data, all the pairs are comparable. Censoring makes some comparisons impossible and by not using them (original option) bias is introduced. We know that censoring up to the largest event time can effectively be dealt with the procedure presented by Uno et al. (2011). For discussion of censoring after the largest event time see Kejžar et al. (2016).

In this note, we limit ourselves to no censoring to leave the equations simple.

Time-dependent covariates are commonly present in studies of survival and time-varying effects are often found during the analysis. Harrell's c-index was defined for constant effects and covariates, meaning that the predictions are made at time 0. If covariates and/or covariate effects change in time, predictions have to change. This means that the original c-index cannot be used.

There were quite some proposals to include time dependency, but before discussing (some of) them, we first introduce some notation.

The variables of interest are the true survival time T_i and the predicted survival time T_i^* and we denote their observed values by t_i and t_i^*. T^* is usually a function of predictor variables X. The concordance parameter can be expressed as

$$\mathscr{C} = P(T_i^* < T_j^* | T_i < T_j).$$

Its estimator is

$$c = \frac{\sum_{i=1}^n \sum_{j=1}^n I(t_i < t_j) I(t_i^* < t_j^*)}{\sum_{i=1}^n \sum_{j=1}^n I(t_i < t_j)}, \tag{1.2}$$

where I is the indicator function. This is of course the same as Equation (1.1).

2. Generalisations

There is little doubt that what is meant by the notion of a generalisation is the following: a generalised c-index would give the same value as the c-index if the prediction model had no time-varying effects (coefficients) and time-dependent covariates. Therefore the

Definition 1. A statistic is a generalisation of the c-index if it is equal to the c-index under the following conditions:

1. the prediction model has no time-dependent covariates
2. the prediction model has no time-varying effects.

In the literature, the *time-dependent c-index* is used in two different ways that are important to distinguish. Time-dependency can mean (i) that c-index is computed at different (final) *time-points* and (ii) that time-dependent predictions are included in the computation of c-index. The true time-dependent statistic should be able to consider both. C-index modifications that address only (i) are computed at a certain time t that divides the data set in two groups: those that had the event by t and the rest, and evaluate the measure at that time point (Chambless & Diao, 2006; Gerds et al., 2013).

To include time-dependent predictions (point (ii)) into the computation of the statistic, the statistic has to be flexible enough to compare model predictions and observations at each event time. To achieve this, one has to partition the data set in two at every single event time, calculate (partial) concordance there and summarise over the whole time span.

2.1. Different functions for T^*

To illustrate different approaches to generalisations we have chosen three, published in statistical literature.

Antolini et al. (2005) propose a generalisation of c-index defined by

$$\widehat{\mathscr{C}_1} = \frac{\sum_{t_{(k)}} P\left[\hat{S}(t_i|X_i(t)) < \hat{S}(t_i|X_j(t)) \wedge t_i < t_j \wedge t_i = t_{(k)}\right]}{\sum_{t_{(k)}} P\left[t_i < t_j \wedge t_i = t_{(k)}\right]}$$

where $t_{(k)}$ denotes the ordered k-th time of events. Indexes i and j denote all comparable units at $t_{(k)}$, where also $t_{(k)} = t_i$ holds. P corresponds to sample probability and $\hat{S}(y|X)$ is a function for predicted survival probability at time y with covariates equal to X. For time-dependent covariates $X(t)$, t denotes the time instants where there are covariate variations. Note that predictions may be obtained by any type of survival model. At time $t_{(k)}$ only units with observed time greater or equal to $t_{(k)}$ are compared and they are concordant if their predicted survivals at $t_{(k)}$ are in line with that (i.e., the larger the predicted survival, the longer the actual time of the event).

Antolini's equation can be rewritten in a way to sum over all units (instead of ordered event times):

$$\widehat{\mathscr{C}_1} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} I(t_i < t_j)\mathbf{I}\left[\hat{\mathbf{S}}(\mathbf{t_i}|\mathbf{X_i}(\mathbf{t})) < \hat{\mathbf{S}}(\mathbf{t_i}|\mathbf{X_j}(\mathbf{t}))\right]}{\sum_{i=1}^{n}\sum_{j=1}^{n} I(t_i < t_j)}.$$

This equation resembles Equation (1.2) with the only distinction in the term in bold which denotes the function for T^* used to compute the rank.

Heagerty and Zheng (2005) in their paper review the extensions of diagnostic accuracy measures (sensitivity and specificity) to survival data. They propose the *incident/dynamic* definition which accounts for the multiple contributions that a unit i can make to the model at different event times.

It is shown (Heagerty & Zheng, 2005) that the AUC_t (the area under the receiver operating characteristic at given t) equals $P(T_i^* < T_j^* | T_i < T_j \wedge T_i = t)$ which is the concordance parameter at time t. The function for T_i^* in Heagerty and Zheng (2005) is taken to be the prognostic index from a survival regression model, hence a monotone function of the hazard.

The weighted average of the $AUC(t)$ is (as shown in the Appendix of Heagerty & Zheng, 2005) the overall c-index:

$$\widehat{\mathscr{C}_2} = \sum_{t_{(k)}} AUC_{t_{(k)}} \cdot w^{t_{(n)}}(t_{(k)}).$$

$w^{t_{(n)}}(t)$ denotes the weight for each AUC_t, computed from the observed overall survival and rescaled to sum to 1 at $t_{(n)}$. If the prognostic index varies in time, the c-index accounts for that.

Stare et al. (2011) proposed a measure of explained variation (R_E for ranks explained) for survival data for time-dependent covariates and/or time-varying effects. Its estimator reduces to c-index when there is no time-dependency, and uses IPC weights for the correction of bias. The idea of the measure is to rank model-based intensity estimates at each distinct event time and summarise the extent to which the model matches the ranking of the data.

The estimate of the measure with no bias correction (and no ties) is

$$\widehat{R}_E = \frac{\sum_{t_i}(r_{i,\text{null}} - r_{i,\text{model}})}{\sum_{t_i}(r_{i,\text{null}} - r_{i,\text{perfect}})}$$

The concurrent ranks for unit i at time t_i are defined as

$$\begin{aligned} r_{i,\text{null}} &= \frac{|\mathscr{R}_{t_i}|+1}{2} && \text{average rank} \\ r_{i,\text{perfect}} &= 1 && \text{best rank} \\ r_{i,\text{model}} &= |\mathscr{R}_{t_i}| - \sum_{j=1}^{n} I(t_i < t_j) I\left[h(t_i|X_i(t)) > h(t_i|X_j(t))\right] \end{aligned}$$

where $h(t)$ represents the predicted hazard at time t. $\mathscr{R}_t$ denotes the risk set at t. A close look reveals that $r_{i,\text{model}}$ is computed as the maximal rank minus the number of all concordant pairs at t_i. Imputing the rank expressions into the equation of $\widehat{R}_E$ we get

$$\widehat{R}_E = -1 + 2 \cdot \frac{\sum_{t_i}\left(\sum_{j=1}^{n} I(t_i < t_j) I\left[h(t_i|X_i(t)) > h(t_i|X_j(t))\right]\right)}{\sum_{t_i}(|\mathscr{R}_{t_i}| - 1)}.$$

The denominator of the second term represents the number of all comparable pairs for each t_i. The numerator is twice the number of concordant pairs for each t_i, therefore the whole term equals time-dependent c-index $\widehat{R}_E = 2\widehat{\mathscr{C}}_3 - 1$. The time-dependent c-index, in this case, is of the form

$$\widehat{\mathscr{C}}_3 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} I(t_i < t_j)\mathbf{I}\left[\mathbf{h}(\mathbf{t_i}|\mathbf{X_i}(\mathbf{t})) > \mathbf{h}(\mathbf{t_i}|\mathbf{X_i}(\mathbf{t}))\right]}{\sum_{i=1}^{n}\sum_{j=1}^{n} I(t_i < t_j)},$$

which resembles Equation (1.2). The rank of T^* is computed by the use of hazard function (the term in bold), similarly as proposed in the paper of Heagerty and Zheng (2005). Note that the link between the measure estimator $\widehat{R}_E$ and c-index $\widehat{\mathscr{C}}_3$ is the same as between Kendall's τ and the original c-index.

Harrell's c-index is usually computed for Cox regression models with proportional hazards where the function for predicting T^* is usually the hazard $e^{X^\top\beta}$ or the prognostic index, the monotonic transformation of hazard. In that setting $X_i^\top\beta < X_j^\top\beta$ corresponds to $S(t|X_i) > S(t|X_j)$ (Antolini et al., 2005). However with time-dependent covariates and/or time-varying effects that does not hold anymore. Survival function $S(t)$, as well as cumulative distribution function $1 - S(t)$, are cumulative measures ($S(t) = P(T > t)$), and hazard function is an instantaneous measure of risk. If hazard modifies, one detects that immediately and its relative change is larger than in survival function. In survival the whole history is also accumulated and that makes relative changes of two time points smaller with time.

The definition of time-dependent c-index should compute partial concordances for specific time points (event times) and a natural choice for the function for predicting T^* there is hazard (or its monotonic transformation). We are aware that models that do not model hazards, such as random forests, might be used to model event history data; in that case, the estimation of the second term of Equation (1.2) is not that transparent and the results do not apply.

2.2. Simulation results

We found no expressions for population value and the variance of the described estimators $\widehat{\mathscr{C}_1}$ and $\widehat{\mathscr{C}_2}$, hence to demonstrate the difference of time-dependent c-indices defined in the previous section we simulate a simple time-dependent data set (only time-varying effect is considered in this short illustration). Times to events are simulated from exponential distribution with one binary covariate. The effect of covariate changes through time. It is positive ($\beta_1 = 2$) from the beginning till $t = 0.2$ and negative afterwards ($\beta_2 = -0.5$). There is no censoring. In real life this simulation (with possibly different time units and rate of decrease in survival) could mimic the survival after major surgery in cancer patients. Figure 1 shows the respective Kaplan-Meier curves for a simulated data set ($n = 300$). We note that the curves on the left graph do not cross. We fit the Cox model with fixed effect and calculate the original c-index, and the Cox model with time-varying effect and calculate $\widehat{\mathscr{C}_1}$ (*survival*), $\widehat{\mathscr{C}_2}$ (*AUC*), and $\widehat{\mathscr{C}_3}$ (*hazard*). We find that the value of $\widehat{\mathscr{C}_1}$ is the same as the *original* since the ranking of units throughout the time does not change. $\widehat{\mathscr{C}_2}$ and $\widehat{\mathscr{C}_3}$ are close to each other, but different from the original, meaning that they react to change in the effect even though the curves do not cross. The crossing of curves in the right graph affects $\widehat{\mathscr{C}_1}$, however much smaller than on $\widehat{\mathscr{C}_2}$ or $\widehat{\mathscr{C}_3}$. We note that the data for the two graphs were generated with the same underlying model.

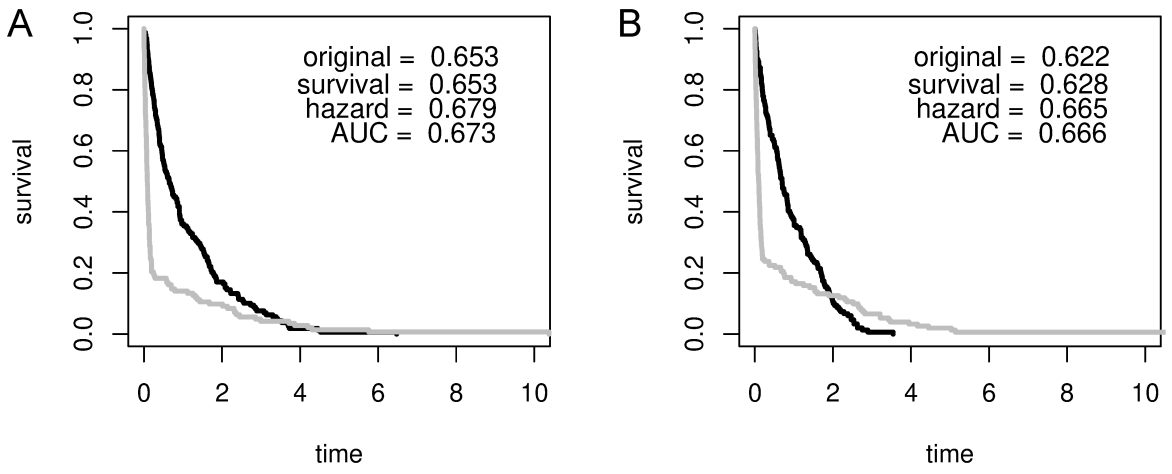

Figure 1: Two data sets ($n = 300$) for survival data where effect changes from positive to negative at 0.2. (A) Kaplan-Meier survival curves for controls (black) and cases (gray) do not cross; (B) do cross.

Figure 2 shows distributions of the respective four indices for 10^5 simulations. They confirm the message from Figure 1.

A note on censoring. The R function from the paper of Stare et al. (2011) for $\widehat{\mathscr{C}_3}$ corrects for bias, as does implicitly also the calculation of AUC ($\widehat{\mathscr{C}_2}$) in the R statistical package `risksetROC` (Heagerty, 2012). In the widely used statistical package `survival` (Therneau, 2021) c-index is automatically calculated when running a Cox model and yields identical results to the hazard estimator $\widehat{\mathscr{C}_3}$ when using time-dependent models in the case of no

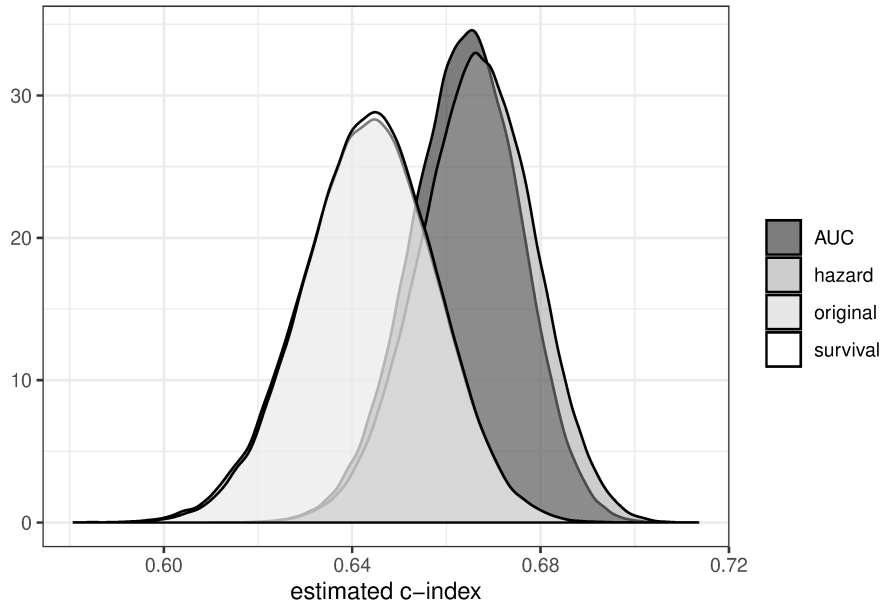

Figure 2: 10^5 realisations of the four differently defined c-indices.

censoring. When censoring is included, it does not correct for bias (see right panel of Figure 3 where empirical densities for the c-indices are presented, same simulation model as in Figure 2 with 35% censoring).

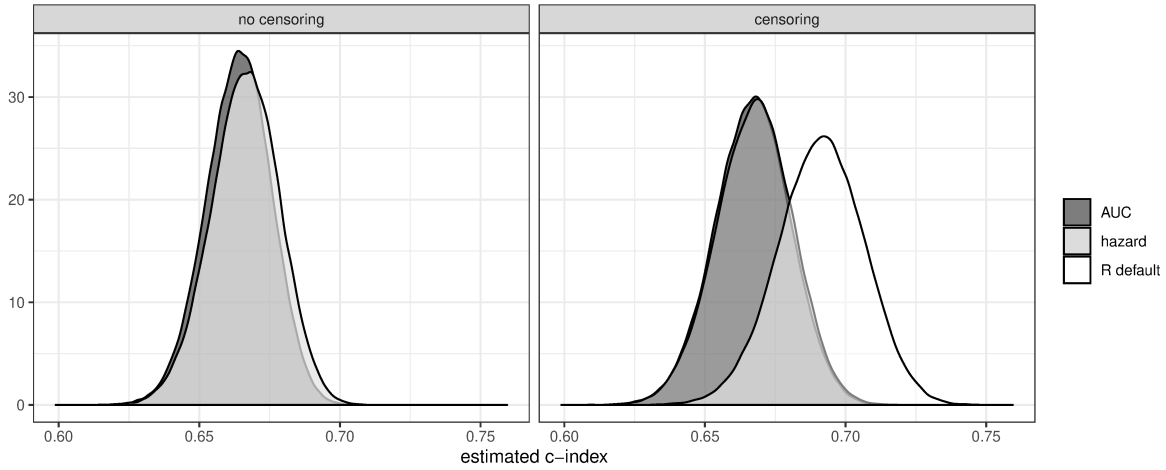

Figure 3: 10^5 realisations of the c-indices $\widehat{\mathscr{C}_2}$ (AUC), $\widehat{\mathscr{C}_3}$ (hazard), and the one implemented in `survival` package in R (default) and their sensitivity to censoring (35%, $n = 10^5$).

3. Conclusions

Generalising Harrell's c-index means changing the predictions at each time point at which either the values of some covariates change or the effects (coefficients) of some covariates change. The predictions can be based on (i) survival functions (or cumulative hazards) or (ii) instantaneous hazards. The second option allows for changes to be observed more rapidly and thus recommended. Since survival function is a cumulative measure of risk (i.e. hazard), an additional solution would be to find a one-to-one correspondence between the generalisations which would make all options recommended.

Acknowledgments

This work was supported by the Slovenian Research Agency (Methodology for data analysis in medical sciences, P3–0154).

References

Acion, L., Peterson, J. J., Temple, S., & Arndt, S. (2006). Probabilistic index: An intuitive non-parametric approach to measuring the size of treatment effects. *Statistics in Medicine*, *25*(4), 591–602. https://doi.org/10.1002/sim.2256

Antolini, L., Boracchi, P., & Biganzoli, E. (2005). A time-dependent discrimination index for survival data. *Statistics in Medicine*, *24*(24), 3927–3944. https://doi.org/10.1002/sim.2427

Chambless, L. E., & Diao, G. (2006). Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statistics in Medicine*, *25*(20), 3474–3486. https://doi.org/10.1002/sim.2299

Gerds, T. A., Kattan, M. W., Schumacher, M., & Yu, C. (2013). Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring. *Statistics in Medicine*, *32*(13), 2173–2184. https://doi.org/10.1002/sim.5681

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Harrell, J., Frank E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *JAMA: The Journal of the American Medical Association*, *247*(18), 2543–2546. https://doi.org/10.1001/jama.1982.03320430047030

Heagerty, P. J. (2012). *Risksetroc: Riskset ROC curve estimation from censored survival data* (Version 1.0.4) [Computer software]. The Comprehensive R Archive Network. https://cran.r-project.org/package=risksetROC

Heagerty, P. J., & Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics*, *61*(1), 92–105. https://doi.org/10.1111/j.0006-341X.2005.030814.x

Kejžar, N., Maucort-Boulch, D., & Stare, J. (2016). A note on bias of measures of explained variation for survival data. *Statistics in Medicine*, *35*(6), 877–882. https://doi.org/10.1002/sim.6749

Kremers, W. K. (2007). *Concordance for survival time data: Fixed and time-dependent covariates and possible ties in predictor and tim*. Mayo Foundation. https://www.mayo.edu/research/documents/biostat-80pdf/doc-10027891

Lehmann, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. *The Annals of Mathematical Statistics*, *22*(2), 165–179. https://doi.org/10.1214/aoms/1177729639

Stare, J., Pohar Perme, M., & Henderson, R. (2011). A measure of explained variation for event history data. *Biometrics*, *67*(3), 750–759. https://doi.org/10.1111/j.1541-0420.2010.01526.x

Therneau, T. M. (2021). *Survival: Survival analysis* (Version 3.2-13) [Computer software]. The Comprehensive R Archive Network. https://cran.r-project.org/package=survival

Uno, H., Cai, T., Pencina, M. J., D'Agostino, R. B., & Wei, L. J. (2011). On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*, *30*(10), 1105–1117. https://doi.org/10.1002/sim.4154