Comparison of Survival Forests in Analyzing First Birth Interval

Mahsa Saadati[1], Arezoo Bagheri[1*]

1. Associate professor of National Population Studies & Comprehensive Management Institute, Tehran, Iran.

Abstract

Background and objectives: Application of statistical machine learning methods such as ensemble based approaches in survival analysis has been received considerable interest over the past decades in time-to-event data sets. One of these practical methods is survival forests which have been developed in a variety of contexts due to their high precision, non-parametric and non-linear nature. This article aims to evaluate the performance of survival forests by comparing them with Cox-proportional hazards (CPH) model in studying first birth interval (FBI).

Methods: A cross sectional study in 2017 was conducted by the stratified random sampling and a structured questionnaire to gather the information of 610, 15-49-year-old married women in Tehran. Considering some influential covariates on FBI, random survival forest (RSF) and conditional inference forest (CIF) were constructed by bootstrap sampling method (1000 trees) using R-language packages. Then, the best model is used to identify important predictors of FBI by variable importance (VIMP) and minimal depth measures.

Results: According to prediction accuracy results by out-of-bag (OOB) C-index and integrated Brier score (IBS), RSF outperforms CPH and CIF in analyzing FBI (C-index of 0.754 for RSF vs 0.688 for CIF and 0.524 for CPH and IBS of 0.076 for RSF vs 0.086 for CIF and 0.107 for CPH). Woman's age was the most important predictor on FBI.

Conclusion: Applying suitable method in analyzing FBI assures the results which be used for making policies to overcome decrement in total fertility rate.

Introduction

In biostatistics, an active area of research is survival analysis which concentrates on time-to event outcomes that are typically censored (1-3). Since new progress in data learning technology has made high dimensional data available to researchers in recent years, newfound challenges in analyzing survival data is appeared. In this situation, traditional survival analysis methods such as Cox proportional hazard (CPH) regression that are quite useful due to the simple interpretations of the covariate effects and readily inferences, become inadequate (1, 4). The drawbacks of these models could be summarized as forcing a specific link between the covariates and the response, specifying interactions between covariates by the analyst, and in practice, inferenceing is often made after trying many models. Another vital issue is violating the proportional hazards (PH) assumption in high-dimensional time-to-event data. To overcome these problems, non-parametric and flexible methods such as survival trees and tree ensembles have been developed (5-8). They offer great flexibility and can automatically detect certain types of interactions without the need to specify them beforehand. Survival forests as an ensemble non-parametric method are very powerful predictive tools comparing to a single tree which can be extracted from many fitted survival trees (7, 8). More recently, random survival forest (RSF) (9, 10), as an extension of Breiman's random forest techniques (11), constructed by bagging of classification trees for survival data, has been proposed as an alternative method for improving survival prediction and variable selection. RSF has several advantages of being completely data driven, independent of model assumptions and robust to outliers in the covariate space (12), and representing a suitable tool for exploratory analysis where survival prior information is limited. In case of high dimensional data, RSF do not face univariate regression limitations such as over-fitting, unreliable estimation of regression coefficients, and inflated standard errors (13). RSF have been criticized for the bias that results from favoring covariates with many split-points and hence conditional inference forest (CIF) for time to event data have been suggested (14). However, the two models are comparable in predictive performance on time-to-event data sets with categorical covariates that are binary in nature. The superiority in performance of the CIF model is likely due to the way it handles the split variable and the split point selection especially in the presence of covariates with many split-points. Though, it is a challenging issue to result which of them is the best in general (14).

To the best of the authors' knowledge, RSF and CIF have not been applied frequently in the demography field of study and little references are available in this area (14, 15). However, there are many studies in medical science that have been used this approach (16-20).

In this paper, one of the aims is reviewing basic ideas of survival forest algorithms in the following section. The performance of CPH and survival forests in analyzing the FBI using the data of "effects of socio-economic rationality dimensions on childbearing behavior in Tehran" survey in 2017 (21) and R-language packages (random Forest SRC, party and pec) are compared. At the end, the brief discussion and conclusion are also presented.

Materials and Methods

The main aim of this article is to compare KM, CPH, RSF and CIF survival analysis methods in modeling FBI according to the prediction error curves, Harrell's C-index and Integrated Brier score (IBS). After selecting the best model, the most important covariates will be introduced by considering minimal depth and VIMP rankings measures.

Denote the true survival time by a continuous random variable Y and the censoring time by a continuous variable C. Let T = min(Y, C) be the observation time and δ =I (Y$\leq$C) be the right censoring indicator, where δ = 1 represents the observation is an event and δ = 0 represents the observation is censored. Let X = (x1, x2, $\cdots$, xp) denote a p-dimensional covariate vector and D be the dataset containing n independent and identically distributed observations sampled from (X, T, δ) (i=1,…,n), namely D (12). CPH is generally described as:

$$h(Y) = h_0(Y)\exp(b_1x_1 + ...+ b_p\,x_p) \quad (1)$$

where h(Y) denotes the hazard given the values X and the respective survival time (t). The term h0(Y) is called the baseline hazard that is the hazard for the respective individual when the values of all the covariates are equal to zero. The use of the Cox model allows us to determine the relationship between the hazard rate and covariates without specifying baseline hazard function. This model assumes that the hazard function for an individual depends on the values of the covariates and the baseline hazard. As a result, given two individuals with particular values for the covariates, the ratio of the estimated hazards over time will be constant (PH assumption). When this important hypothesis is not confirmed, the model is not effective and its results are not reliable.

Several survival forest methods have been proposed in the literature (22, 23). The main output of RSF method (9), which is the most closely related to the original random forest method (11), is an estimated cumulative hazard function computed by averaging the Nelson-Aalen cumulative hazard function of each tree. The RSF algorithm consists of the following steps:

(a) Drawing L bootstrap samples or in-bag (IB) observations with replacement from a training data set of size n (about two-thirds of the original data). The remaining out-of-bag (OOB) observations (30 % of data) will not appear in the bootstrap sample.

(b) For each bootstrap sample, grow a full size survival tree based on a certain splitting criterion without pruning. At each internal node, randomly select candidate covariates out of all p covariates. Candidate covariates, which minimize the risk within the nodes or maximize the separation between the nodes, are used for splitting. Stop growing until a certain stopping condition is met (e.g., when the number of observations within a terminal node is less than a preset value or when the node becomes pure). By default, the number of candidate covariates is and the log-rank statistic is the splitting rule.

(c) For each tree, a cumulative hazard function (CHF) is calculated. For a particular terminal node h at time point t, CHF is estimated by the Nelson-Aalen estimator as:

$$\hat{H}_h(t) = \sum_{L,h \le t} \frac{d_{L,h}}{Y_{L,h}} \quad (2)$$

where $d_{(L,h)}$ and $Y_{(L,h)}$ are the number of interested events and individuals at risk at time point $t_{l,h}$,respectively. Hence, all

observations within the same node have the same CHF.

(d) To predict the cumulative hazard of a new observation x, average over all CHFs from all the L trees to obtain the ensemble CHF of the forest:

$$\hat{H}_E(t|x) = \frac{1}{L}\sum_{i=1}^{L} \hat{H}_i(t|x) \tag{3}$$

where $\hat{H}_i(t|x)$ denotes the CHF of the tree grown from the ith bootstrap sample.

The CIF or cforests, for survival analysis (24) is another fully non-parametric, utilizes a weighted Kaplan-Meier estimate based on all subjects from the training data at x for prediction. The approach of tree aggregation in CIF is different from RSF. CIF put more weight on terminal nodes where there are a large number of subjects at risk. In contrast, RSF use equal weights on all terminal nodes. However, it is difficult to say which formula may be better in general (25). The CIF algorithm consists of the following steps (14):

(a) For case weights w, the global null hypothesis of independence between any of the p covariates and the response variable is tested. If this hypothesis cannot be rejected, It will be stopped, otherwise the jth covariate Xj* with strongest association to T is selected.

(b) Select a set A* ⊂ Xj* in order to split Xj* into two disjoint sets. The weights wα and wβ determine the two subgroups with wα,i =wiI (Xj*,i ∈ A*) and wβ,i = wiI (Xj*,i A*) for all i =1, 2, . . . , n.

(c) Recursively steps (a) and (b) with modified case weights wα and wβ, respectively is repeated.

There are four different performance statistics that are defined as following:

Harrell's C-index: Prediction accuracy for survival forests was assessed by Harrell's C-index using OOB data. The C-index is calculated using the following steps:

All possible pairs of subjects are formed.

Permissible pairs by eliminating those pairs whose shortest survival time is censored, by eliminating pairs (i,j) if $T_i=T_j$,and both experienced the interested event, are considered.

For each permissible pairs where $T_i=T_j$, count 1 if the shorter survival time has high risk predicted. Count 0.5 if risk predicted is tied; for each permissible pair, where $T_i=T_j$ and both experienced the interested event, count 1 if risk predicted is tied otherwise count 0.5. For each permissible pair where $T_i=T_j$, but at least one is censored, count 1 if the event has right risk predicted, otherwise count 0.5. Let concordance denote the sum over all permissible pairs.

C-index= concordance/permissible.

A value of 0.5 for C-index is not better than random guessing and a value of 1 denotes full-discriminative ability.

Integrated Brier score (IBS): The Brier score at time t is given by:

$$\tag{4}$$

$$\P BS(t) = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\left(0-\hat{s}(t|x_i)\right)^2}{\hat{G}(t_i)}I(T_i \leq t, \sigma_i = 1) + \frac{\left(1-\hat{s}(t|x_i)\right)^2}{\hat{G}(t_i)}I(T_i > t)\right\}$$

where $\hat{G}(t)=P(C_i>t)$ denote the Kaplan-Meier estimate of the censoring survival function (26, 27). The prediction error curve is gotten by calculating of Brier score across

the times. In addition, IBS that is cumulative prediction error curves over time is given:

(5)

$$\P IBS = \frac{1}{\max(t_i)} \int_0^{\max(t_i)} BS(t)dt$$

Lower values of IBS indicate better predictive performances.

Variable Importance (VIMP): it measures the increase (or decrease) in prediction error for the forest ensemble when a variable is randomly noised up (11). VIMP is the difference in OOB prediction error before and after permutation, so a large VIMP value indicates that mis-specification detracts from the predictive accuracy in the forest. VIMP close to zero indicates the variable contributes nothing to predictive accuracy, and negative values indicate the predictive accuracy improves when the variable is misspecified (28).

Minimal depth: Ishwaran et al. (9) represented minimal depth as a new high dimensional variable selection method based on a tree concept. This measure differs from traditional method for variable selection in RSF which has been based on VIMP (11). The complex nature of the VIMP calculation has made it difficult to study (13). In contrast, minimal depth uses rigorous theory for selecting variables as well as comprehensive methodology for regularizing forests and inspects the forest construction instead of ranking covariates (9, 20, 29). Minimal depth assumes that covariates with high impact on the prediction are those that most frequently split nodes nearest to the root node, where they partition the largest samples of the population. Within each tree, node levels are numbered based on their relative distance to the root of the tree (with the root at 0). It

measures important risk factors by averaging the depth of the first split for each variable over all trees within the forest. The assumption in the metric is that smaller minimal depth values indicate the variable separates large groups of observations, and therefore has a large impact on the forest prediction. Ishwaran et al. (29) also derives a simple optimistic threshold rule uses the mean of the minimal depth distribution, classifying variables with minimal depth lower than this threshold as important in forest prediction.

In the following section, the rationale of applying survival forests comparing to CPH is illustrated by analyzing women's FBI data according to covariates of their age, age at first marriage, educational level, partner educational level, activity, region, house ownership, kinship, ideal birth interval, race, partner race, expenditure, marriage time attitude, migration status, sex preference of the introduced data set.
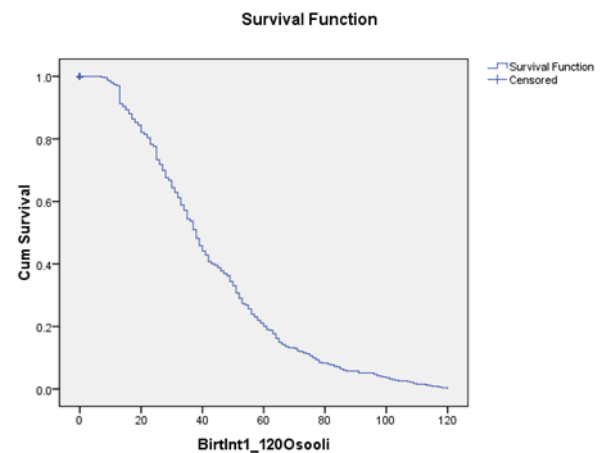
Results

In a cross-sectional study, a structured questionnaire was reviewed from 610 married Iranian women aged 15-49 years selected by multi-stage stratified random sampling from different regions of Tehran province in 2017 (21). The structured questionnaire collected demographic, fertility history and childbearing attitudinal factors. The questionnaire's validity was confirmed by 10 demographers and sociologists, and its reliability was at least 0.771 for questionnaire's factors by Cronbach's Alpha. Since there were not any interventions or treatment in this study, and the aim of the study was explained to the respondents before interviewing, therefore there is no requirement for ethical code.

23.1 percentages of women were childless. Mean of women's age and their first marriage age were 35.22±7.91 and 22.61±4.67 years, respectively. Most of them were in age 30-39 (43.4%), unemployed (68.0%), non-migrant (86.1%), Fars (56.7%), renter (50.3%), and with no sex preference (81.6%). 86.7% of women had diploma and above degree, 58.0% had ideal birth interval greater than 3 years, 57.4 % married for the first time in age 20-29 years, 72.5% had non-family husband, 68.7% lived in developing and middle developing regions, and 59.3 % had less than 2 million Tomans expenditure in a month. 81.6 percentages of their husband had diploma and above educational level and 55.7 percentages of them were Fars. Kaplan-Meier (KM) mean of the FBI was 38±1.06 months; only women's age (p-value=0.000), educational level (p-value=0.000), partner educational level (p-value=0.001), activity (p-value=0.014) and region (0.020) had

significant effect on the FBI based on log-rank test.

KM survival estimates are computed for women's FBI and their survival curve is shown in Figure (1). As this figure displays, most of the women's FBI (about 88%) were less than 5 years.


Figure 1. Kaplan-Meier survival curve of first birth interval

To investigate effects of all covariates on FBI simultaneously, CPH model was applied and the results has been presented in Table (1). It should be noted that PH hypothesis for all covariates were checked through correlating the corresponding set of scaled Schoenfeld residuals with time, to test for independence between residuals and time. The results have also been presented in Table (1). PH assumption is supported by a non-significant relationship between residuals and time, and refused by a significant relationship.

**Table 1.** Cox model for first birth interval

Variable		β	Hazard Ratio(HR)	Std. Error	P-value	Chisq	P-value
			Cox model			**PH assumption test**	
Age		0.018	1.018	0.008	**0.021**[*]	2.150	0.143
Age at First Marriage	<20(ref)						
	20-29	-0.117	0.890	0.116	0.314	11.100	**0.001**[*]
	>30	0.063	1.065	0.208	0.762	2.460	0.117
Educational level	Under-Diploma (ref)						
	Diploma and above	-0.244	0.783	0.173	0.458	0.583	0.445
Partner Educational level	Under-Diploma(ref)						
	Diploma and above	-0.176	0.839	0.144	0.222	0.311	0.577
Activity	Unemployed(ref)						
	Employed	-0.166	0.847	0.120	0.166	0.037	0.847
Region	Developing and middle-developing (ref)						
	Developed and more developed	-0.148	0.862	0.114	0.193	1.850	0.174
House Ownership	Renter (ref)						
	Owner	0.052	1.054	0.110	0.634	0.001	0.976
	Others	0.023	1.023	0.175	0.896	0.298	0.585
Kinship	Family (ref)						
	Non-family	0.319	1.375	0.115	0.006[**]	20.700	0.000[**]
Ideal Birth Interval	<=3 (ref)						
	>3	-0.091	0.913	0.098	0.353	0.637	0.425
Race	Fars(ref)						
	Azari	0.010	1.010	0.144	0.945	0.040	0.841
	Others	-0.075	0.928	0.182	0.683	0.000	0.993
Partner Race	Fars(ref)						
	Azari	0.236	1.266	0.144	0.102	0.184	0.668
	Others	0.172	1.188	0.182	0.344	5.580	0.018[*]
Expenditure	Less than 2 Million(ref)						
	2-3.5 Million	0.000	1.000	0.116	0.998	0.891	0.345
	3.5 Million and more	0.132	1.141	0.179	0.459	0.605	0.436
Marriage Time Attitude	Sooner(ref)						
	Later	-0.198	0.820	0.176	0.259	4.050	0.044[*]
	On-time	0.021	1.021	0.163	0.899	0.603	0.437
Migration Status	Migrant(ref)						
	Non-migrant	0.071	1.074	0.147	0.628	0.200	0.655
Sex Preference	Yes(ref)						
	No	0.110	1.116	0.128	0.390	0.923	0.337

The first approach for comparing fitted models is to see how each model's prediction error fluctuates as time goes on. The prediction error curves for KM, CPH, RSF and CIF are presented in Figure (2).
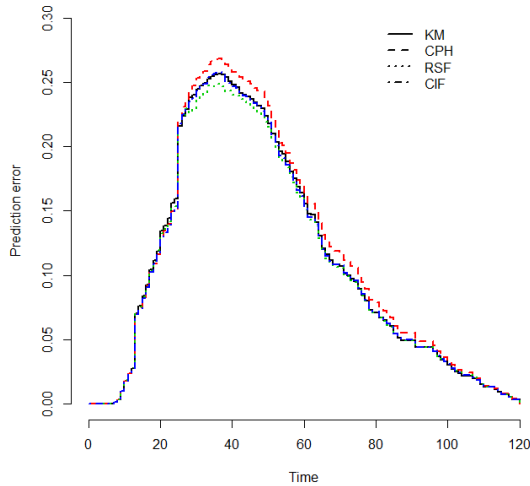
Figure 2. Comparison of prediction error curves

The second measure for comparing survival analysis methods makes use of each model's OOB C-index over time. Figure (3) displays box plot of OOB C-index for CPH, RSF, and CIF.
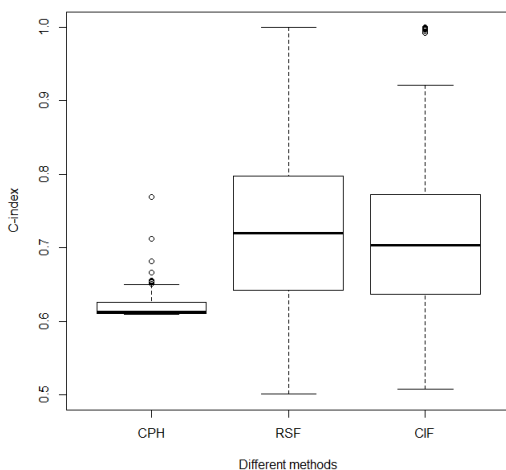

Figure 3. Box plot of C-index comparison

The third comparison method is IBS which is cumulative prediction error over time. IBS for RSF, CPH, and CIF is equal to 0.076, 0.107and 0.086, respectively.

Figure (4) left panel shows the error rate for the RSF log-rank model as a function of the number of trees and right panel indicates the VIMP for predictors with left (negative) and right (positive) hand side bars.
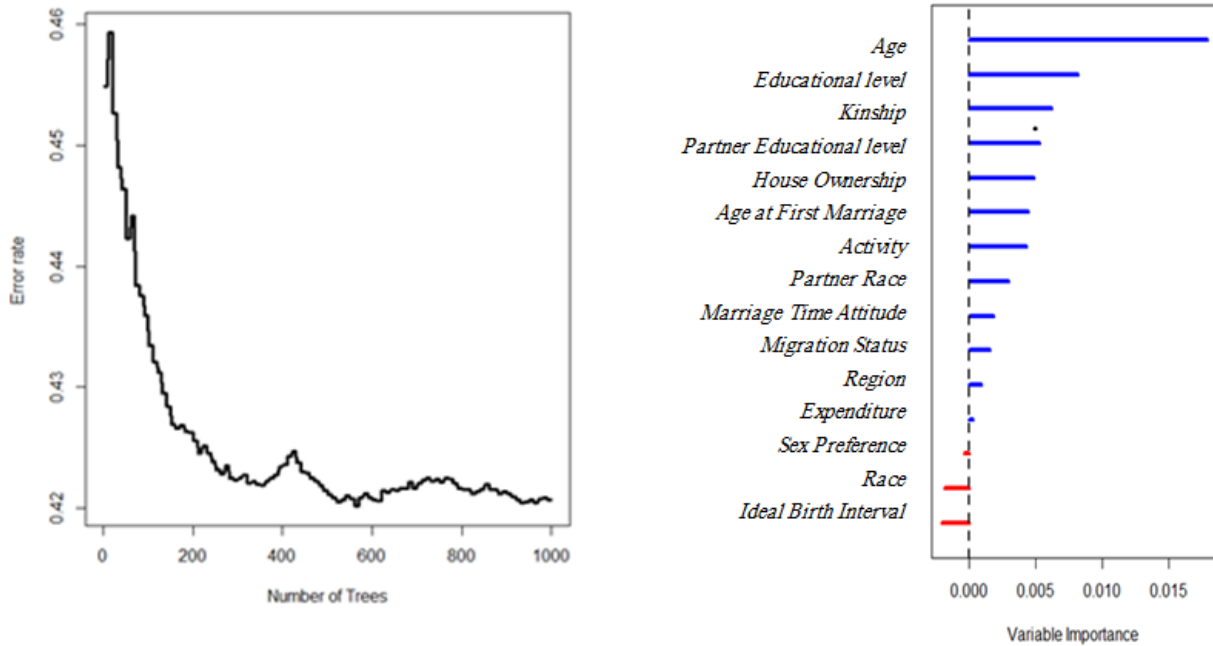
Figure 4. The prediction error rate (left panel) and VIMP (right panel) for random survival forest first birth interval

Figure (5) displays the comparison of minimal depth and VIMP rankings measures, simultaneously. Since these measures use different criteria, it is expected that the variable ranking to be somewhat different. Horizontal and vertical axes indicate VIMP and minimal depth with their threshold lines. The points along the bisector line shows where the measures of VIMP and minimal depth are in agreement. Points above this line are ranked higher by VIMP than by minimal depth, indicating the variables are more sensitive to misspecification. Those below the line have a higher minimal depth ranking, indicating they are better at dividing large portions of the population.
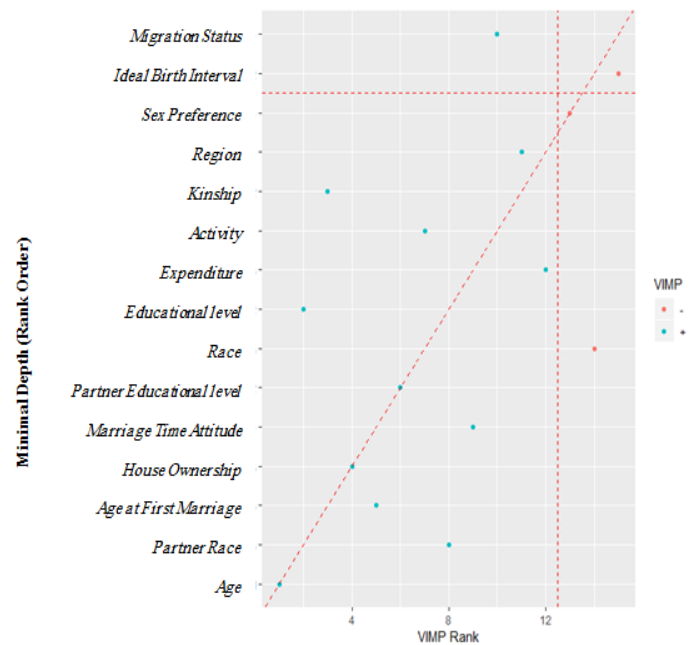
Figure 5. Comparing Minimal Depth and VIMP rankings

Discussion

According to the results of Table (1), only women's age (p-value=0.021) and kinship to their husband (p-value=0.006) had significant effects on their FBI. Moreover, the PH assumption test is just statistically significant for covariates of age at first marriage (p-value=0.001), kinship (p-value=0.000), partner race (p-value=0.018) and marriage time attitude (p-value=0.044). Therefore, we cannot assume the PH. In this context, it is unrealistic to expect the reported Cox coefficients to be satisfactory indicators of the actual covariate effects on FBI.

According to Figure (2), all four curves are practically the same and the only real differences among prediction error curves are in around 30-50 months. The maximum and minimum prediction errors in this time belong to CPH and RSF, respectively.

In Figure (3), it is obvious that RSF clearly outperforms the other models in prediction performance of FBI survival according to its largest median of C-index (0.754) comparing to the CPH (0.524) and CIF (0.688) medians.

Another important measure for confirming the best method to analyze FBI is IBS. The smaller value is, the better accuracy of prediction is achieved. According to the results, the smallest IBS belongs to RSF (0.076) which outperforms the other two methods.

Therefore according to the results of prediction performance measures, the best model to predict FBI is RSF. The left panel of Figure (4) results in the prediction error rate of 42.11 % for RSF. But the right depicts the VIMP for all 15 covariates. From the plot, it could be indicated that among all covariates only women's sex preference, race, and ideal birth interval had negative VIMP values on childlessness survival time.

In Figure (5), the minimal depth threshold is 4.931. Among 15 selected covariates in studying FBI, both VIMP and minimal depth consider women's age, house ownership and partner educational level as important and sex preferences as non-important covariates. However, considering just VIMP measure results in non-importance of women's race and ideal birth interval covariates and designating just minimal depth ends to the conclusion of non-importance of migration status covariate.

Conclusion

Survival analysis methods are a mainstay of the biostatistics field that are recently finding growing use in other disciplines including demography (1-3). An extensively used tool in survival analysis is CPH model with the main PH assumption which may not be a good approximation to reality (4). In contrast, survival forests are non-parametric methods and have the flexibility to model survivor curves that are of dissimilar shapes for contrasting groups of subjects (12-14).

The main purpose of this paper was to compare survival forests with CPH in studying FBI, and determine which variables contribute to the prediction accuracy. The main finding of this study was that prediction errors estimated RSF clearly outperforms the other strategies according to the prediction performance measures. Thus, analyzing researches by CPH should be done with cautious. The finding of this article gives the same result as some researches (16-18, 25, 30, 31). Some studies like as (15, 22) suggested that CIF are superior in predictive performance to RSF on time-to-event datasets with polytomous covariates. However, the two models are comparable in predictive performance with categorical covariates that are binary in nature. The superiority in

performance of the CIF model is likely due to the way it handles the split variable and the split point selection especially in the presence of covariates with many split-points. It is therefore important for researchers to select the best survival forest model to analyze any time-to-event dataset based on the nature of its covariates (32).

Acknowledgements

Funding source(s)

Conflict of interest

All the authors declare that no financial or non-financial conflicts of interest referred to the subject matter or materials discussed in the article exist.

References

1. Cox DR. Analysis of survival data: Routledge; 2018. [DOI:10.1201/9781315137438]

2. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data: Springer Science & Business Media; 2006.

3. Elashoff R, Li N. Joint modeling of longitudinal and time-to-event data: CRC Press; 2016. [DOI:10.1201/9781315374871]

4. David CR. Regression models and life tables (with discussion). Journal of the Royal Statistical Society. 1972;34(2):187-220. [DOI:10.1111/j.2517-6161.1972.tb00899.x]

5. Gordon L, Olshen RA. Tree-structured survival analysis. Cancer treatment reports. 1985;69(10):1065-9.

6. Hothorn T, Bühlmann P, Dudoit S, Molinaro A, Van Der Laan MJ. Survival ensembles. Biostatistics. 2005;7(3):355-73. [DOI:10.1093/biostatistics/kxj011]

7. Zhang H, Singer BH. Recursive partitioning and applications: Springer Science & Business Media; 2010. [DOI:10.1007/978-1-4419-6824-1]

8. Bou-Hamad I, Larocque D, Ben-Ameur H. A review of survival trees. Statistics Surveys. 2011;5:44-71. [DOI:10.1214/09-SS047]

9. Ishwaran H, Kogalur UB. Random survival forests for R. R news. 2007;7(2):25-31.

10. Ishwaran H, Lu M. Random survival forests. Wiley StatsRef: Statistics Reference Online. 2008:1-13. [DOI:10.1002/9781118445112.stat08188]

11. Breiman L. Random forests. Machine learning. 2001;45(1):5-32. [DOI:10.1023/A:1010933404324]

12. LeBlanc M, Crowley J. Relative risk trees for censored survival data. Biometrics. 1992;June 1:411-25. [DOI:10.2307/2532300]

13. Dietrich S, Floegel A, Troll M, Kühn T, Rathmann W, Peters A, et al. Random Survival Forest in practice: a method for modelling complex metabolomics data in time to event analysis. International journal of epidemiology. 2016;45(5):1406-20. [DOI:10.1093/ije/dyw145]

14. Nasejje JB, Mwambi H, Dheda K, Lesosky M. A comparison of the conditional inference survival forest model to random survival forests based on a simulation study as well as on two applications with time-to-event data. BMC medical research methodology. 2017;17(1):115-42. [DOI:10.1186/s12874-017-0383-8]

15. Nasejje JB, Mwambi H. Application of random survival forests in understanding the determinants of under-five child mortality in Uganda in the presence of covariates that satisfy the proportional and non-proportional hazards assumption. BMC research notes. 2017;10(1):459-77. [DOI:10.1186/s13104-017-2775-6]

16. Adham D, Abbasgholizadeh N, Abazari M. Prognostic factors for survival in patients with gastric cancer using a random survival forest. Asian Pacific journal of cancer prevention: APJCP. 2017;18(1):129-34.

17. Miao F, Cai Y-P, Zhang Y-T, Li C-Y, editors. Is random survival forest an alternative to Cox proportional model on predicting cardiovascular disease? 6TH European conference of the international federation for medical and biological engineering; 2015: Springer. [DOI:10.1007/978-3-319-11128-5_184]

18. Yosefian I, Mosa Farkhani E, Baneshi MR. Application of random forest survival models to increase generalizability of decision trees: a case study in acute myocardial infarction. Computational and mathematical methods in medicine. 2015:1-7. [DOI:10.1155/2015/576413]

19. Wang H, Li G. A Selective Review on Random Survival Forests for High Dimensional Data. Quantitative bio-science. 2017;36(2):85. [DOI:10.22283/qbs.2017.36.2.85]

20. Ishwaran H, Kogalur UB, Chen X, Minn AJ. Random survival forests for high-dimensional data. Statistical Analysis and Data Mining: The ASA Data Science Journal. 2011;4(1):115-32. [DOI:10.1002/sam.10103]

21. Abdolahi A. Effects of socio-economic rationality dimensions on childbearing behavior in Tehran. National Population Studies & Comprehensive Management Institute,; 2017.

22. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical statistics. 2006;15(3):651-74. [DOI:10.1198/106186006X133933]

23. Krętowska M, editor Random forest of dipolar trees for survival prediction. International Conference on Artificial Intelligence and Soft Computing; 2006: Springer. [DOI:10.1007/11785231_95]

24. Hothorn T, Lausen B, Benner A, Radespiel-Tröger M. Bagging survival trees. Statistics in medicine. 2004;23(1):77-91. [DOI:10.1002/sim.1593]

25. Mogensen UB, Ishwaran H, Gerds TA. Evaluating random forests for survival analysis using prediction error curves. Journal of statistical software. 2012;50(11):1-23. [DOI:10.18637/jss.v050.i11]

26. Gerds TA, Schumacher M. Consistent estimation of the expected Brier score in general survival models with right-censored event times. Biometrical Journal. 2006;48(6):1029-40. [DOI:10.1002/bimj.200610301]

27. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. Statistics in medicine. 1999;18(17-18):2529-45. https://doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2529::AID-SIM274>3.0.CO;2-5 [DOI:10.1002/(SICI)1097-0258(19990915/30)18:17/183.0.CO;2-5]

28. Ehrlinger J. ggRandomForests: Exploring random forest survival. arXiv preprint arXiv:161208974. 2016.

29. Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High-dimensional variable selection for survival data. Journal of the American Statistical Association. 2010;105(489):205-17. [DOI:10.1198/jasa.2009.tm08622]

30. Weathers B. Comparision of Survival Curves Between Cox Proportional Hazards, Random Forests, and Conditional Inference Forests in Survival Analysis. 2017:1-35.

31. Hsich E, Gorodeski EZ, Blackstone EH, Ishwaran H, Lauer MS. Identifying important risk factors for survival in patient with systolic heart failure using random survival forests. Circulation: Cardiovascular Quality and Outcomes. 2011;4(1):39-45. [DOI:10.1161/CIRCOUTCOMES.110.939371]

32. Das A, Abdel-Aty M, Pande A. Using conditional inference forests to identify the factors affecting crash severity on arterial corridors. Journal of safety research. 2009;40(4):317-27. [DOI:10.1016/j.jsr.2009.05.003]