Survival Trees and Direct Adjusted Survival Curves—Prediction of Survival Probabilities

Wioletta Grzenda

Abstract There are two main methods of constructing survival trees. The first method is based on measures of heterogeneity of survival functions in individual nodes, while the second method uses intranodal differentiation determined by the likelihood function or the partial likelihood function as a criterion for the division. The Kaplan–Meier estimator is used to estimate the survival curve for individuals located at terminal nodes. The prediction of survival probabilities for a given individual based on the thus obtained conditional Kaplan–Meier curve does not consider their characteristics, which were omitted in the construction of the divisions and in some cases may lead to conclusions that are too general. Then, the solution may be to use the direct adjusted survival curve, for the construction of which all explanatory variables included in the Cox model are used. In this article, we compare these two survival prediction methods, paying attention to the limitations and advantages of each. The empirical analysis was carried out with the use of data from the 2018 Labor Force Survey for Poland. The economic activity of women around retirement age was examined.

Keywords Survival trees · Direct adjusted survival curves · Kaplan–Meier curves · Women's employment

1 Introduction

Prediction in the survival analysis is based on the estimation of the survival function or the cumulative hazard function. Non-parametric methods are most often used to determine the estimators of these functions. The two basic ones are the Kaplan–Meier method (Kaplan and Meier 1958) and the Nelson-Aalen method (Aalen 1978; Nelson 1972). One important limitation of these methods is that covariates which may affect the duration of an individual in a definite state are not considered when constructing the survival function estimators. Therefore, these functions are most

W. Grzenda (✉)
SGH Warsaw School of Economics, Warsaw, Poland
e-mail: wgrzend@sgh.waw.pl

K. Jajuga et al. (eds.), *Modern Classification and Data Analysis*,
Studies in Classification, Data Analysis, and Knowledge Organization,
https://doi.org/10.1007/978-3-031-10190-8_3

often determined for groups of individuals designated by the categories of a specific qualitative variable. In this approach, the other characteristics of the individuals are not considered at all. The simultaneous influence of many factors on the duration of an individual in a definite state can be accounted for using parametric and semi-parametric models (Blossfeld and Rohwer 1995).

Machine learning methods are increasingly being used to describe the relationships between the characteristics of an individual and their duration in a definite state; among them the most popular are the survival trees (Zhou and McArdle 2015). Tree construction for censored survival data cannot be immediately transferred from classification or regression tree algorithms due to the lack of a natural measure of homogeneity within the node. The concept of survival trees was first described in (Gordon and Olshen 1985). The authors of this work point out the simplicity of the proposed methods, while noting that their assumptions concern only the traceability of conditional duration distributions under the condition of certain explanatory variables.

The purpose of recursive splitting methods used to construct trees for censored survival data is to divide the population into homogeneous groups with regard to the duration of individuals in a definite state. There are two main methods of constructing survival trees based on the maximization of heterogeneity between nodes (Molinaro et al. 2004). The first one is based on measures of heterogeneity between nodes (Ciampi et al. 1986; De Rose and Pallara 1997), among them the log-rank test statistic is the most frequently used (Peto and Peto 1972; Klein and Moeschberger 2006). The second method is based on the CART algorithm (the Classification and Regression Tree (CART) algorithm) (Breiman et al. 1984) and the likelihood or partial likelihood function (Davis and Anderson 1989; Al-Nachawati et al. 2010).

Both types of survival tree models allow not only to define groups of individuals like each other in terms of survival time but can also be used to determine the probability of duration of individuals in a definite state. The prediction of survival probabilities of an individual obtained in this way depends on their individual characteristics. However, in the case of trees, whose structure is based on the comparison of survival curves determined with the Kaplan–Meier estimator, it is not possible to include time-dependent variables in their construction. This restriction does not apply to trees based on Cox regression (Cox 1972; Cox and Oakes 1984). They enable the prediction of survival probabilities for individuals characterized by both time-independent and time-dependent covariates. However, this approach also has a certain limitation, as it requires verification of the assumption of proportional hazards.

The prediction of survival probabilities based on the survival trees is performed by determining the survival function for the terminal nodes (leaves). If the distributions of explanatory variables omitted in the construction of a given leaf are not uniform in the selected groups, such approach may lead to erroneous conclusions. Then, a better solution may be to use the semi-parametric Cox model and the direct adjusted survival curves for prediction of survival probabilities (Chang et al. 1982; Gail and Byar 1986; Zhang et al. 2007). In the case of prediction of survival probabilities based on the adjusted survival function, all the characteristics of the individuals included in the model are considered, but this approach also requires verification of the assumption

of proportional hazards. This study compares the results obtained with the use of two types of survival trees and the adjusted survival function. The aim of the study was to indicate the advantages and limitations of these prediction methods in the survival analysis as well as to analyze the discrepancy in the results obtained with these methods. Moreover, to increase the accuracy of prediction, we extended the algorithm proposed of Al-Nachawati et al. (2010) by introducing multiple splits in the process of construction of survival trees. The duration of employment of women aged 51–70 was considered as a reference problem.

2 Theoretical Backgrounds

In the survival analysis, the basic function used to describe the duration of an individual in a definite state is the survival function:

$$S(t) = P(t < T) \tag{1}$$

where T denotes the variable describing the time until the event occurs. The most common estimator of the survival function, also used in this work, is the Kaplan–Meier estimator (Kaplan and Meier, 1958). It is given by the following formula:

$$\widehat{S}(t) = \prod_{j:t_j<t} \left(1 - \frac{d_j}{n_j}\right) \tag{2}$$

where d_j denotes the number of events that occurred at t_j and n_j is the number of individuals at risk of the event until t_j.

The comparison of survival curves estimated by the Kaplan–Meier method is the basis for the construction of many survival trees (Ciampi et al. 1986; De Rose and Pallara 1997). The construction of trees of this type is based on the algorithm of recursive division of the multidimensional feature space into disjoint subsets due to differences in the survival time of the considered individuals. The log-rank test, called the Mantel-Cox test (Klein and Moeschberger 2006), is most often used to compare the survival distribution in the selected groups. The division of the feature space is made into two disjoint areas in a recursive manner, until the whole space is divided into many areas differentiated by the survival distribution. A detailed description of the tree construction algorithm of this type can be found in the work of LeBlanc and Crowley (1993). The prediction of survival probabilities with the use of such model of the survival tree consists in determining the survival function for each terminal node using the Kaplan–Meier method.

There is also the second type of survival trees (Davis and Anderson 1989; Al-Nachawati et al. 2010), the construction of which is based on Cox regression (Cox 1972; Cox and Oakes 1984). Let $\mathbf{x} = [x_1, \ldots, x_k]^T$ denote a vector of covariates

and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_k]$ a vector of estimated model parameters. Then, for the Cox proportional hazards model, the hazard is given by the following formula:

$$h(t|\mathbf{x}) = h_0(t)\exp(\mathbf{x}\boldsymbol{\beta}) \tag{3}$$

where $h_0(t)$ denotes baseline hazard.

Let us assume that for each jth, $j = 1, \ldots, N$ individual t_j signifies their survival time and v_j is a censoring variable, which takes the value 1, for event-affected individuals and 0 for censorship. Let x_{jq} for $q = 1, \ldots, p$ denote qth question for jth individual. Then the hazard function $h_j(t)$ for jth individual and qth question x_{jq} is given by the following formula:

$$h_j(t) = h_0(t)\exp\left(b_q x_{jq}\right) \tag{4}$$

where b_q denotes the unknown coefficients. Each explanatory variable can be used to split the parent node with a certain probability. This probability is determined by testing the global null hypothesis $\boldsymbol{\beta} = 0$ in a semi-parametric regression model. The construction of the tree in this study was carried out according to the algorithm proposed by Al-Nachawati et al. (2010), with the difference that multiple, not only binary splits were considered.

The algorithm for constructing a survival tree based on Cox regression can be represented as follows:

Step 1. Estimate the Cox proportional hazards regression model using the forward technique, considering all covariates available in the data set. Find the variable with the smallest p-value.

Step 2. The first split is made based on the variable identified in step 1, but for a categorical variable with more than two levels, a multiple split is used, and the levels are not linked.

Step 3. For each subgroup of individuals located in each of the nodes created in step 2, the Cox model is estimated. A forward technique determines a covariate, based on which the next split of each node is made.

Step 4. Repeat step 3 for the obtained nodes as long as there is a statistically significant covariate determined by the forward technique or the previously defined stop criterion, taking into account, for example, the size of the tree, is not met.

To derive the prediction of survival probabilities based on this model of the survival tree, similar to the previously considered tree, the survival function is determined for each terminal node by the Kaplan–Meier method.

An alternative to the Kaplan–Meier method of determining the survival function is the Cox regression approach (Cox 1972; Cox and Oakes 1984), which enables the determination of the survival function separately for each group of individuals characterized by a specific set of their features. In the semi-parametric Cox model, the formula for the survival function is as follows:

$$S(t) = [S_0(t)]^{\exp(\mathbf{x}\boldsymbol{\beta})} \tag{5}$$

where $S_0(t)$ is the baseline survival function corresponding to the baseline hazard $h_0(t)$. The baseline survival function S_0 can be presented using the cumulative hazard function H_0 as follows:

$$S_0(t) = \exp(-H_0(t)) \tag{6}$$

where $H_0(t) = \int_0^t h_0(u)du$, $t \geq 0$. The estimator of the survival function $S(t)$ presented in this way has the following form:

$$\widehat{S}(t) = \left[\widehat{S}_0(t)\right]^{\exp\left(\mathbf{x}\widehat{\boldsymbol{\beta}}\right)} \tag{7}$$

where $\widehat{\boldsymbol{\beta}} = \left[\widehat{\beta}_1, \ldots, \widehat{\beta}_k\right]$ denotes the estimator of the parameter vector $\boldsymbol{\beta}$ and $\widehat{S}_0$ is the estimator of a baseline survival function, which is given by the following formula:

$$\widehat{S}_0(t) = \prod_{u|t_{(u)}<t} \left(1 - \frac{d_u}{\sum_{l \in R(t_{(u)})} exp\left(\boldsymbol{x}_l \widehat{\boldsymbol{\beta}}\right)}\right) \tag{8}$$

where d_u, $u = 1, 2, \ldots, m$ is the number of observations, for which the event occurred at the moment $t_{(u)}, u = 1, 2, \ldots, m$, and $R\left(t_{(u)}\right), u = 1, 2, \ldots, m$, denotes a hazard set. The hazard set includes all individuals for which the survival or censoring time is greater than t_u.

Let j now denote an individual belonging to the ith group, then the observed values for this individual can be described by $\left\{t_{ij}, v_{ij}, \mathbf{x}_{ij}\right\}, i = 1, 2, \ldots, K$, $j = 1, 2, \ldots, n_i$, where t_{ij} is the observed time, $v_{ij} = 0$, when censoring occurs and $v_{ij} = 1$ otherwise, and $\mathbf{x}_{ij}$ denotes the covariates vector. Then, the survival function at the moment t, for an individual from the i-th group with values of variables **x,** has the following form (Chang et al. 1982; Gail and Byar 1986; Zhang et al. 2007):

$$\widehat{S}_i(t; \mathbf{x}) = \exp\left\{-\widehat{H}_{0i}(t)\exp\left(\mathbf{x}\widehat{\boldsymbol{\beta}}\right)\right\} \tag{9}$$

The general formula for the direct adjusted survival curve is.

$$\widehat{S}_i(t) = \frac{1}{n}\sum_{l=1}^{n} \exp\left\{-\widehat{H}_{0i}(t)\exp\left(\mathbf{x}_l\widehat{\boldsymbol{\beta}}\right)\right\} \tag{10}$$

where $n = \sum_{i=1}^{K} n_i$.

In this study, the obtained survival function was compared to the survival functions obtained by the Kaplan–Meier method with the use of two types of survival trees. Moreover, with the use of all the methods presented in this section, it is possible to determine the probability of a definite event occurring for the examined individuals and those on which the model was not trained.

3 Empirical Examples

3.1 Data

The study used a data set from the Labor Force Survey (LFS). The research sample consisted of women who were examined for two consecutive quarters of 2018 (samples 75–77 and 79–81). Women aged 51–70 who had ever worked after 2011 were selected for the study. There were 9,540 women who met these criteria.

In survival analysis, the dependent variable is time T, i.e., the variable representing the waiting time until the occurrence of an event. In our study, T denotes the time until employment termination. More precisely, at the time of the study, 60.22% of women were still working—in the study they were considered as censored individuals. For these women, the time was calculated as the number of months from 2011 or from starting work, if it was started after 2011, until the second survey. For women who were no longer working at the time of the study, the time was calculated as the number of months from 2011 or from the moment of starting work, if it was started after 2011, until the moment they stopped working. For these women, the event occurred. Moreover, considering the statutory retirement age for women in Poland in the study, which is 60 years, we defined the *Age_group* variable to distinguish women who have not yet reached the retirement age and those who have already reached that age. Table 1 presents the other characteristics of the women, which were considered in the construction of the survival trees and the direct adjusted survival curves. In addition, in the table, apart from the variable name, its abbreviated name used in the survival trees is given in parentheses.

3.2 Survival Trees

In the first stage of the research, a survival tree with a maximum depth of 4 was constructed based on the data set presented in point 3.1. and using R software. The tree was constructed with the ctree function available in the partykit package. The resulting tree is shown in Fig. 1. Based on the obtained results, it can be concluded that the greatest impact on the termination of the employment relationship of the studied women was reaching retirement age, education, and the type of employment.

Table 1 Sample characteristics

Variable	Description	Levels	Proportion [%]
Age_group (Age)	Age group of women at the time of the survey	1 = from 51 to 59 years old	53.44
		2 = from 60 to 70 years old	46.56
Education (Edu)	Level of education	1 = higher	23.82
		2 = post-secondary or secondary	36.24
		3 = basic vocational or primary school	39.94
Marital_status (Mar)	Marital status	0 = unmarried, a widower, a widow, separated or divorced	27.41
		1 = married	72.59
Place_residence (Res)	Class of place of residence during the survey	1 = city of 100 thousand residents and more	37.51
		2 = city to100 thousand residents	30.41
		3 = rural areas	32.08
Employment (Emp)	Type of employment	1 = salaried employee	79.88
		2 = self-employed or helping family member	20.12
Institutions (Ins)	Institution (company) that is a place of work	1 = no information	20.12
		2 = private	39.48
		3 = public	40.40
Elderly_person (Eld)	The presence of elderly person over 75 years old in the household	0 = no	91.43
		1 = yes	8.57

In the second stage of this study, with the use of SAS software, the semi-parametric Cox model was estimated using the forward technique for all characteristics presented in Table 1, previously verifying the assumption of proportional hazards. The estimation results of this model are included in Table 2, while the summary of the forward selection in Table 3. As a result of the forward technique, it was found that the covariate that has the greatest impact on economic activity is the age of women (Table 3). Therefore, it was the first covariate that was used to split the data. Consequently, the observations in the root were divided into two sets. Two separate Cox proportional hazard models were then estimated using a forward technique for these two received sets of observations using all available women characteristics, except for the age variable, again selecting the most important variable for each of the received sets. For both nodes this is education. Since this covariate has three levels, the observations in each of these two nodes were divided into three sets, and then on each of the extracted sets, the Cox proportional hazard model was re-estimated using a forward technique to determine the covariates against which subsequent divisions

Fig. 1 The decision tree based on the log-rank test (Tree 1)

were made until the tree reached a given depth 4. As a result of such splits, a tree was obtained, as shown in Fig. 2. For each node, the number of observations for which the event occurred (1) and censored (0) was given.

For the construction of the second tree, all the covariates available in the data set were used; however, the covariates that had the greatest impact on the duration of

Table 2 Estimated parameters, standard error, p-value, and hazard ratio for covariates in the model

Covariate	Parameter estimate	Standard error	*p*-value	Hazard ratio
Age group of women at the time of the survey (ref. from 60 to 70 years old)				
From 51 to 59 years old	−1.7662	0.0399	<0.0001	0.171
Level of education (ref. basic vocational or primary school)				
Higher	−0.9449	0.0545	<0.0001	0.389
Post-secondary or secondary	−0.3284	0.0369	<0.0001	0.720
Class of place of residence during the survey (ref. rural areas)				
City of 100 thousand residents and more	−0.1068	0.0445	0.0163	0.899
City to 100 thousand residents	−0.0197	0.0421	0.6400	0.981
The presence of elderly person over 75 years old in the household (ref. yes)				
No	−0.2007	0.0627	0.0014	0.818
Institution (company) that is a place of work (ref. public)				
No information	−0.2537	0.0509	<0.0001	0.776
Private	0.2874	0.0382	<0.0001	1.333

Table 3 Summary of forward selection

Step	Covariate	Number included	Chi-square	Pr > ChiSq
1	Age group of women at the time of the survey	1	2509.7365	<0.0001
2	Level of education	2	428.4950	<0.0001
3	Institution (company) that is a place of work	3	139.9721	<0.0001
4	The presence of elderly person over 75 years old in the household	4	10.5880	0.0011
5	Class of place of residence during the survey	5	6.8367	0.0328

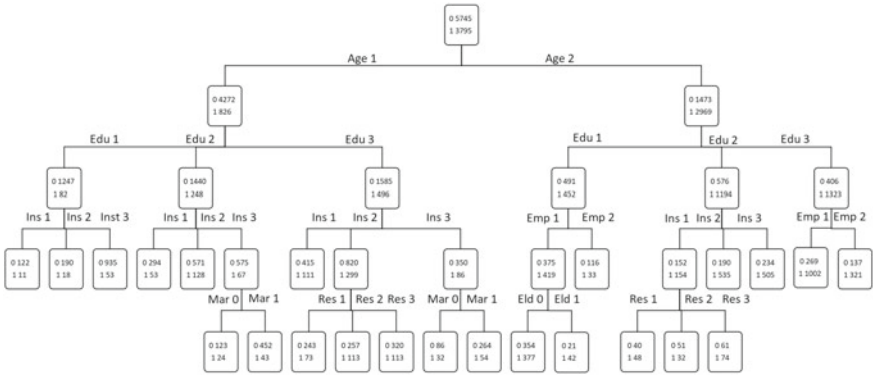

Fig. 2 The decision tree based on Cox regression (Tree 2)

employment of the studied women were the covariates describing their age, education, and the type of employment. These are the same covariates that were used to construct the first tree.

3.3 Survival Function Estimation

The presented models of survival trees make it possible to distinguish groups of women for whom there are differences due to the duration of the employment relationship; however, in line with the purpose of this chapter, the focus was on comparing the prediction of survival probabilities obtained with the use of both survival trees using the direct adjusted survival curve. This paper presents the estimates of the survival function using the Kaplan–Meier method based on the results obtained from models of survival trees for women with the most and least frequent characteristics in the data set. The obtained functions were compared to the direct adjusted survival curve. This curve is the averaged survival function for the individuals that occurred

in the data set under consideration. This means that the survival curves are averaged, not the values of the variables (Zhang et al. 2007).

Figure 3 shows the estimates of the survival curves for women with the most common characteristics in the data set, i.e., women who simultaneously met the following conditions: they had not yet reached retirement age, had basic vocational or primary school education, were married, lived in a city with over 100,000 residents, worked as salaried employees in a public institution (company), and there were no persons over 75 in their households. Based on the obtained curves, it can be seen that each of the methods considered in this paper gave similar results. Moreover, it can be concluded that the probability that a woman with such characteristics will not terminate the employment relationship in the next 8 years is approximately 0.75. The curves start to decrease faster after about 7 years, which means that after this period the chances of staying in employment decrease faster and faster.

Figure 4 shows the estimates of the survival curves for women with the least common features in the sample, i.e., women who have already reached retirement age, have higher education, are not married, live in a city up to 100,000 residents, are self-employed or help with family enterprises; moreover, in the households of these women there were people over 75 years of age. It can be seen from the obtained curves that the curves determined by the Kaplan–Meier method based on the models of survival trees are the same. This is because in both models the women under

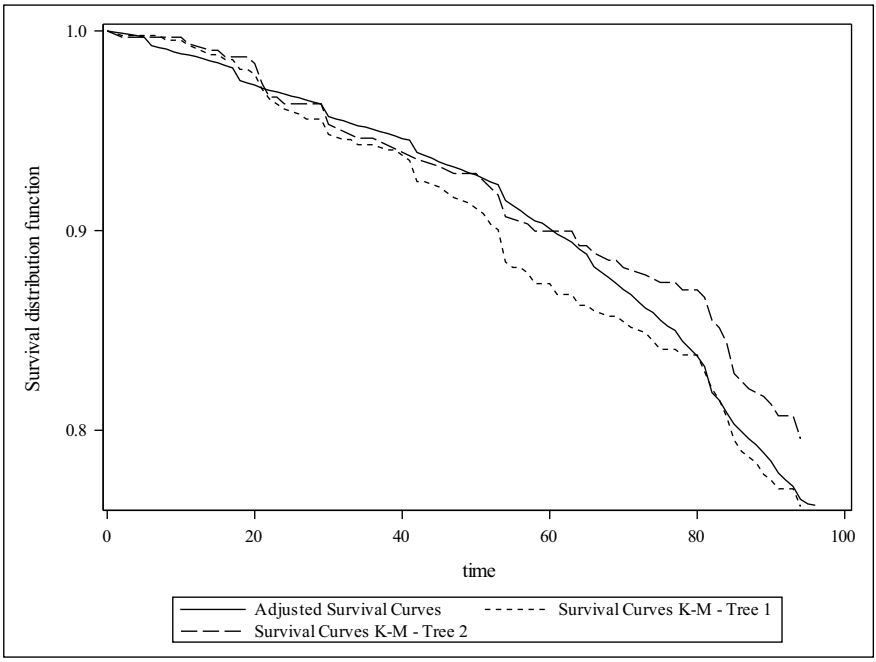

Fig. 3 The comparison of the curve obtained with the Kaplan–Meier method based on trees and the direct adjusted survival curve for women with the most common characteristics in the sample

consideration ended up in the terminal node defined in the same way. Moreover, there are clear differences between the so estimated survival curve and the direct adjusted survival curve, which considers all the characteristics of these women, not only those that influenced their allocation in the terminal node in the models of the trees under consideration. Based on the adjusted survival function, it can be concluded that the probability that a woman with such characteristics will not terminate the employment relationship in the next 8 years is approximately 0.55. However, based on decision tree models, this probability is 0.75. Because these women already reached their retirement age, it can be assumed that the prediction of duration probability in professional activity obtained with the use of the direct adjusted survival curves is more accurate. To verify this assumption, the accuracy of the prediction was assessed using the Schemper and Henderson measure (Schemper and Henderson 2000). Based on the obtained values for this measure, it can be concluded that considering the probability of the duration of economic activity of all the characteristics of these women may improve its accuracy.

4 Summary and Conclusions

This study compares the prediction of survival probabilities methods used in the analysis of survival. With the use of these methods, the probability of employment duration of women approaching or exceeding the retirement age was predicted, considering their individual characteristics. Moreover, based on the results obtained from two types of survival trees, the determinants of economic activity of older women were identified. Our contribution to the research on the economic activity of elderly women in Poland is the identification of the hierarchy of factors that stimulate or limit their staying in the labor market. It was shown that the greatest impact on the professional activity of these women, apart from age, had education and type of work. On the other hand, professional decisions were less influenced by such characteristic as marital status, place of residence, and the presence of a person aged 75 + in the household. Considering the large impact of education on the termination of employment by older women, to keep these women in the labor market, it is worth paying attention to raising their qualifications, or even changing them. On the other hand, the large impact on the employment of women of variables related to the type of work performed and the employment institution may indicate their expectations regarding working conditions are particularly important.

The idea behind the construction of survival trees is to group individuals according to their survival time (LeBlanc and Crowley 1993). In this study, the use of the survival trees to identify the characteristics that define homogeneous groups of women in terms of their duration in employment as well as the variables that most affect this time. However, it has been shown that in the case of prediction of survival probabilities, in some cases, it is better to use the direct adjusted survival curves than survival trees. In addition, Cox regression-based methods enable the construction of a predictive model with both time-independent and time-dependent covariates.

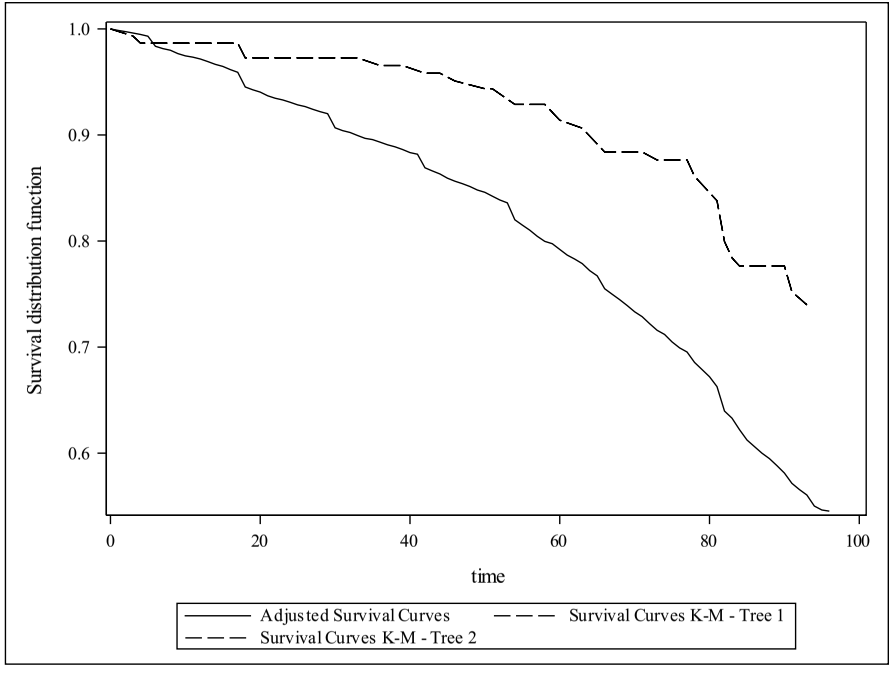

Fig. 4 Comparison of the curve obtained with the Kaplan–Meier method based on survival trees and the direct adjusted survival curves for women with the least common characteristics in the sample

A significant limitation of machine learning methods adapted to censored data is the problem with assessing their accuracy in prediction of survival probabilities. In the case of survival trees, there is a discussion about assessing the quality of a survival model in terms of predictive accuracy, and none of the measures proposed so far has been widely adopted (Zhou and McArdle 2015). The most frequently used measure of predictive accuracy in survival analysis is the one proposed by Schemper and Henderson (Schemper and Henderson 2000), which is based on the survival function estimated without considering additional characteristics and after considering them. Therefore, it enables the assessment of the accuracy of the prediction of survival probabilities from the point of view of the selection of explanatory variables. In the case of survival trees, the method of selecting explanatory variables results from the selected recursive partitioning algorithm. In the presented study, four variables were used for the construction of the tree with a depth of 4, based on the comparison of Kaplan–Meier survival curves, and in the case of the tree based on Cox regression, all variables.

References

Aalen OO (1978) Nonparametric inference for a family of counting processes. Ann Stat 6:701–726
Al-Nachawati H, Ismail M, Almohisen A (2010) Tree-structured analysis of survival data and its application using SAS software. J King Saud Univ-Sci 22(4):251–255
Blossfeld HP, Rohwer G (1995) Techniques of event history modeling. new approaches to causal analysis. L. Erlbaum, New Jersey
Breiman L, Friedman JH, Olshen R et al (1984) Classification and regression trees. Chapman & Hall, New York
Chang IM, Gelman R, Pagano M (1982) Corrected group prognostic curves and summary statistics. J Chronic Dis 35(8):669–674
Ciampi A, Thiffault J, Nakache JP et al (1986) Stratification by stepwise regression, correspondence analysis and recursive partition: a comparison of three methods of analysis for survival data with covariates. Comput Stat Data Anal 4(3):185–204
Cox DR (1972) Regression models and lifetables. J Roy Stat Soc: Ser B (methodol) 34(2):187–202
Cox DR, Oakes D (1984) Analysis of survival data. Chapman and Hall, London
Davis RB, Anderson JR (1989) Exponential survival trees. Stat Med 8(8):947–961
De Rose A, Pallara A (1997) Survival trees: an alternative non-parametric multivariate technique for life history analysis. Eur J Popul 13(3):223–241
Gail MH, Byar DP (1986) Variance calculations for direct adjusted survival curves, with applications to testing for no treatment effect. Biom J 28(5):587–599
Gordon L, Olshen RA (1985) Tree-structured survival analysis. Cancer Treatment Rep 69(10):1065–1069
Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53:457–481
Klein JP, Moeschberger ML (2006) Survival analysis: techniques for censored and truncated data. Springer Science & Business Media, New York
LeBlanc M, Crowley J (1993) Survival trees by goodness of split. J Am Stat Assoc 88(422):457–467
Molinaro AM, Dudoit S, Van der Laan MJ (2004) Tree-based multivariate regression and density estimation with right-censored data. J Multivar Anal 90(1):154–177
Nelson W (1972) Theory and application of hazard plotting for censored survival data. Biometrics 14:945–966

Peto R, Peto J (1972) Asymptotically efficient rank invariant test procedures. J Roy Stat Soc Ser A (general) 135(2):185–198

Schemper M, Henderson R (2000) Predictive accuracy and explained variation in Cox regression. Biometrics 56(1):249–255

Zhang X, Loberiza FR, Klein JP et al (2007) A SAS macro for estimation of direct adjusted survival curves based on a stratified Cox regression model. Comput Methods Programs Biomed 88(2):95–101

Zhou Y, McArdle JJ (2015) Rationale and applications of survival tree and survival ensemble methods. Psychometrika 80(3):811–833