Application of random survival forest for competing risks in prediction of cumulative incidence function for progression to AIDS

Omid Hamidi [(1)], MahtabTapak [(2)], Jalal Poorolajal [(3)], Payam Amini [(4)], Leili Tapak [(5)]

(1) Department of Science, Hamedan University of Technology, Hamedan, 65155, Iran
(2) Social Security Organization, ALINASAB Hospital, Tabriz, Iran
(3) Department of Epidemiology, School of Public Health and Research Center for Health Sciences, Hamadan University of Medical Sciences, Hamadan, 65175-4171, Iran
(4) Department of Epidemiology and Reproductive Health, Reproductive Epidemiology Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran.
(5) Department of Biostatistics and Epidemiology, School of Public Health and Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan 65175-4171, Iran

CORRESPONDING AUTHOR: Leili Tapak - Assistant Professor of Biostatistics, Department of Biostatistics and Epidemiology, School of Public Health and Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan 65175-4171, Iran - Tel: 0098 81 38371394 - E-mail 1: l.tapak@umsha.ac.ir - E-mail 2: leilitapak@gmail.com - Contact phone number: +98 81 38371394

ABSTRACT

Objective: There has remained a need to better understanding of prognostic factors that affect the survival or risk in patients with human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS), particularly in developing countries. The aim of the present study was to identify the prognostic factors influencing AIDS progression in HIV positive patients in Hamadan province of Iran, using random survival forest in the presence of competing risks (death from causes not related to AIDS). This method considers all interactions between variables and their nonlinear effects.

Methods: A data set of 585 HIV-infected patients extracted from 1997 to 2011 was utilized. The effect of several prognostic factors on cumulative incidence function (probability) of AIDS progression and death were investigated.

Result: The used models indicated that tuberculosis co-infection and gender are two top most important variables for cause specific hazards, while age and gender were the most important in predicting cumulative incidence function for AIDS progression in the presence of competing risks, respectively. The patients with tuberculosis had higher predicted cumulative incidence probability. Predicted cumulative incidence probability of AIDS progression was also higher for mother to child mode of HIV transmission. Moreover, transmission type and gender were two top most important variables for the competing event. Men and those patients with IDUS transmission mode had higher predicted risk compared to others.

Conclusion: Considering nonlinear effects and interaction between variables, age was the most important variable in prediction of cumulative incidence probability of AIDS progression and death.

Key words: HIV/AIDS; data mining; competing risks; survival analysis; random survival forest

INTRODUCTION

Currently, the human immunodeficiency virus (HIV) has remained as the first leading cause of death from infectious diseases and a major public health problem all over the world. Acquired immunodeficiency syndrome (AIDS) is the final and most serious stage of HIV infection and leads to sever damages to the body immune system. Since the advent of the epidemic, about 78 million people have been infected with the HIV virus and about 35 million people have died from AIDS-related disease [1, 2].

Despite the lack of functional cure for HIV infection at present, the advent of antiretroviral treatment (ART) has led to a reduction in the HIV-related mortality and has been helpful for patients to return to a relatively healthy and productive lives [3, 4]. Art is also helpful in slowing down AIDS progression process in an HIV-infected person and prolonging their survival [1]. On the other hand, the presence of prognostic factors such as chronic pathologies associated with immunodeficiency, chronic viral and bacterial infections can complicate treatment of the HIV/AIDS [5, 6]. There have been found evidences that indicate life span can be prolonged and the quality of life can be improved significantly if the levels of HIV remain suppressed [6]. On the contrary, co-infection of HIV with other opportunistic infections especially tuberculosis (TB) may lead to an increase in the risk of mortality among the patients.

There has been observed a declination in the HIV-related mortality, nonetheless better understanding of the prognostic factors affecting the survival of the HIV-positive patients is of great importance to improve the life expectancy of the patients [5]. However, limited studies have been conducted regarding survival of HIV-infected patients especially in developing countries and in the Eastern Mediterranean Region [5-7]. Designing the effective intervention strategies aimed to increase the life expectancy of HIV-infected patients is feasible through collecting reliable information about their survival times and the potential risk factors. In this regard, using appropriate statistical models can be helpful in reliable identifying important prognostic factors and improving prediction accuracy of patients' survival. To achieve this, traditional statistical technique, Cox proportional hazards (PH) model, has been widely used to determine potential risk factors in survival data. However, there are restrictive assumptions including proportionality of hazards and linearity of effects on log hazard function (linearity assumption) [8]. Moreover, the performance of Cox regression is not reliable in the presence of high rate of censoring [9].

Ideally, it would be important to improve the predictive performance of the models identifying potential prognostic factors affecting AIDS progression among HIV positive people via learning theory and data mining techniques for survival time. These models require no restrictive assumptions. Recently, machine learning methods such as tree-based approaches have been developed to deal with right censored survival data as well as the presence of competing risks (patients may die without AIDS progression due to some causes not related to AIDS). The promising performance of these models has been confirmed in different areas [10]. Among them, random survival forests (RSF), a non-parametric tree-based ensemble learning method, can automatically handle the difficulties of the Cox model [8, 10].

The present study aimed to identify prognostic factors affecting AIDS progression as well as to predict cumulative incidence function of AIDS in the presence of competing risks using random survival forest.

MATERIALS AND METHODS

Data set

In the present study, we used a data set (obtained from a retrospective cohort study) contained information of 585 patients with HIV/AIDS and extracted from patients' medical records available at the triangular clinics (TCs), in Hamadan, Iran, from 1997 to 2011. The information of the following variables were available in the data set: age, gender, mode of HIV transmission (injection drug users [IDUs], sexual, mother to child, IDUs/sexual, unknown), co-infection with TB, date of HIV diagnosis, date of progression to AIDS, date of death (if any), cause of death (if known) and receiving combination antiretroviral therapy (cART).

An individual with HIV infection regardless of clinical stage confirmed by laboratory criteria according to country definitions and requirements was called an HIV-infected case [6]. An HIV case and AIDS case are defined as follows respectively [11-13]: 1) an individual whose two sequential enzyme-linked immunosorbent assay tests were positive for HIV antibody followed and confirmed by a Western blot test; 2) an individual with a presumptive or definitive diagnosis of stage 3 or stage 4 condition and/or CD4 count<350 per mm 3of blood in an HIV-infected subject.

The outcome of interest for investigation in this study was the duration of time between diagnosis of HIV infection and progress to AIDS. So, AIDS progression was the event of interest and death from causes not related to AIDS was the competing event, because it precludes AIDS progression. The patients who were lost to follow-up or did not experience any of them were considered as censored. The effect of several prognostic factors on the survival time to AIDS was investigated including gender, age, marital status, mode of HIV transmission and co-infection with TB.

Statistical Method

For an HIV-infected patient, there is a possibility of experiencing death without developing AIDS. This suggests that the HIV/AIDS process should be modeled by statistical

model developed for competing risks data (a situation in which an event (or events) precludes the event of interest). In the present study, the RSF algorithm proposed by [14] for competing risks framework was used. This approach has several advantages as well as the following key features: (a) cumulative incidence function (CIF), which represents the probability that an event of a special type has occurred by time t, can be directly estimated; (b) accurate prediction performance is provided; (c) non-linear effects and interactions are modeled; (d) it can be used for event-specific selection of risk factors; and (f) it is free of model assumptions [14].

A RSF [8] consists of a collection of survival trees that are randomly grown by using an independent bootstrap sample from the learning data using random variable selection (at each node). Similar to the RSF in single point time to event setting, trees in competing risk forests are grown very deeply with many terminal nodes (the ends of the tree), but splitting rules to grow the tree and the estimated values calculated within the terminal nodes used to define the ensemble are different [14]. There are two conceptually different approaches to grow a competing risk forest: (1) to grow separate competing risk trees for each of the events (say J events) in each bootstrap sample and to use event-specific splitting rules to grow the trees; and (2) to grow a single competing risk tree in each bootstrap sample where the splitting rules can be either event-specific, or combine event-specific splitting rules across the events. The later approach is more efficient and sufficient for many tasks and was used in the present study.

The steps of the RSF algorithm for competing risks are as follow [14]:

1. to draw B bootstrap samples from the original data and to exclude about 37% of the data in each sample (out-of-bag or OOB data);
2. to grow a competing risk tree for all bootstrap samples based on randomly selected $M \leq p$ candidate variables at each node of the tree. The candidate variable for splitting each node is the one that maximizes a splitting rule (here there are three possibilities: generalized log-rank test, Gray's test, composite splitting rule);
3. to grow the trees to full size under the constraint that a terminal node should have no less than $n0 > 0$ unique cases;
4. to calculate cumulative incidence functions as well as cumulative cause specific hazards for all J events for each tree, b;
5. to take average all over the B trees for each estimators.

In the present study, two separate RSF competing risk forests were fitted and each forest composed of 1000 competing risk trees (with trees constructed from independent bootstrap samples). Each tree was grown based on about 63% of the data on average and the remaining unused out-of-bag data (37%) was used to calculate out-of-bag cross-validated survival for each patient as well as variable importance (VIMP) and minimal depth measures for each independent variable. Two different survival splitting rules were used by the separate RSF competing risk forests. In the first, trees were grown using a generalized log-rank splitting rule which is most suitable for selecting variables that affect cause specific hazards. In the second RSF, trees were grown using a modified Gray splitting rule which is most suitable for identifying variables directly affecting cumulative incidence function.

To evaluate the performance of RSF for competing risks data analysis, we used cross validation to compare the RSF with cause specific and subdistribution hazards models [15]. The used models were compared integrated Brier score (IBS) criterion [16].

All statistical analyses were performed using the "randomForestSRC" [17], "cmprsk" [18] and "pec" [16] packages from the R software, version 3.3.1 (The R Foundation for Statistical Computing, Vienna, Austria; http://www.r-project.org).

RESULTS

The mean (standard deviation) age of the patients was 32.59 (8.71) years. There were 134 (23%) patients who developed AIDS, 134 (23%) patients died from causes not related to AIDS before developing AIDS and 314 (54%) patients experienced no events (censored). Other patients, who were alive or lost to follow up at the end of the study, were considered as censored. Table 1, shows the characteristics of the patients. Most of the patients were male (88.95%), single (45.42%), aged 25-44 years (78.18%), not co-infected with TB (96.37) and injecting drug users (81.51%). Figure 1 illustrates cumulative incidence function for AIDS and death. As seen, cumulative incidence probability for AIDS progression is lower than the competing event of death.

We applied two random survival forests (using two different splitting rules) for both AIDS progression as an event of interest and death from causes not related to AIDS as a competing event to select variables that affect cause-specific hazards as well as cumulative incidence function directly. Table 2 shows the variable importance (VIMP) and minimal depth values for all used covariates in RSFs that can be used to rank variables. The value of VIMP that is greater than 0.002 is assumed as an effective variable. Moreover, the smaller values of minimal depth indicate better predictiveness of the variables. We used the VIMP and minimal depth to rank all included variables. Larger values of the VIMP were related to the variables with better rank. Two columns of Table 2 correspond to the rank of the used variables in RSF (generalized log rank splitting rule) for both events (death and AIDS) that are constructed according to related VIMP columns. According to this table, the largest VIMP value for the event of interest (AIDS progression in HIV positive

TABLE 1. Characteristics of the study population infected with the HIV virus (n=585) [2].

Variables	Number	Percent
Gender		
Female	515	88.95
Male	64	11.05
Marital status		
Single	263	45.42
Married	223	38.52
Divorced	74	12.78
Widow	19	3.28
Tuberculosis infection		
No	558	96.37
Yes	21	3.63
Antiretroviral therapy		
No	11	8.21
Yes	123	91.79
Modes of HIV transmission		
Injecting drug users	472	81.51
Sexual	72	12.45
Mother to child	9	1.55
Injecting drug users/ Sexual	26	4.49
Age group (year)		
1-24	70	12.03
25-44	455	78.18
45-74	57	9.79

patients) belonged to Co-infection with TB (0.057) which is larger than the cut point of 0.002. So it is the first top rank variable in AIDS progression. Gender, transmission type, age at diagnosis and marital status had also VIMP greater than 0.002, respectively. So, they played an important role in predicting AIDS progression in HIV infected patients as well. In addition, for the competing event (death from causes not related to AIDS), the greatest value of VIMP belonged to transmission variable (0.036). According to the table, co-infection with TB had the lowest VIMP (much smaller than 0.002). On the other hand, based on the minimal depth values, the first top most important variable for both events of AIDS and death was age at diagnosis, while the second top most important variable for AIDS progression and death were gender and transmission type, respectively.

We also analyzed data using cause-specific and subdistribution hazards models. Table 3 (a and b) shows the results. As shown (Table 3 (a)), in the cause specific hazards model co-infection and transmission type variables were statistically significant (P<0.05) for the event of interest and age at diagnosis was significant for the competing event. So, according to the table the cause specific hazard of AIDS progression for an HIV positive patient who has co-infected with TB was 9.32 times larger compared with a patient without TB co-infection (adjusted for other variables). Moreover, according to Table 3 (b), being co-infected with TB and transmission type (mother to child) increases the cumulative incidence of AIDS dramatically. Transmission type and age at diagnosis had also significant positive impact on cumulative incidence probability of the competing event (P<0.05).

Figure 2 and figure 3, display 3-year predicted cumulative incidence function (probability) according to the variables for AIDS progression and death from causes not related to AIDS in the patients, respectively. The highest 3-year predicted cumulative incidence function belongs to the patients co-infected with TB and divorced. Moreover, female and those with mother to child transmission type had higher predicted CIF. As can be seen for age variable, the three-year predicted cumulative incidence probability for AIDS progression decreases as age increases up to 30 and after that it tends to increase with age. For the competing event, according to figure 3, it is expected that the cumulative incidence of death be higher in the patients with IDU transmission type, male and those aged more than 35 be at higher risk of death. As seen for age variable, after 35 point the slope of the graph increases sharply. We also provided prediction error curve to compare performance of the models (Figure 4). As seen in all cases RSF outperformed its classical counterpart model. We also compared the predicted cumulative incidence functions of two events in different level of each covariate. The predicted cumulative incidence probabilities of AIDS progression and death were statistically significant for males, coinfection levels (both levels), marital status (all levels), transmission type (IDUs and mother to child type) and age groups (1-24, 25-44 and 45-70 years old) (P<0.001).

To compare the performance of the used RSF methods with traditional counterparts of cause specific hazards and subdistribution hazards, a cross validation method was applied. So the data were randomly divided into train (70%) and test (30%) sets. We repeated the method 100 times. Table 4, shows the results of the cross-validation over 100 repetitions. A seen in both cause-specific and subdistribution hazards cases, the RSF counterpart showed better performance. For example, in subdistribution hazards case RSF resulted in lower IBS (prediction error: 0.061±0.002 for AIDS and 0.152±0.002 for competing event) compared with Fine and Gray model (prediction error: 0.082±0.003 for AIDS and 0.191±0.004 for competing event).

TABLE 2. Variable importance (VIMP) and minimal depth of the variables used in random survival forest for HIV data

Variables	VIMP		VARIABLES' RANK*		MINIMAL DEPTH		VARIABLES' RANK	
	AIDS	Death	AIDS	Death	AIDS	Death	AIDS	Death
Co-infection (HIV/TB)	0.057	-0.002	1	5	2.858	2.334	5	5
Gender	0.011	0.025	2	2	1.206	1.806	2	4
Transmission	0.010	0.036	3	1	1.551	1.301	3	2
Age at diagnosis	0.005	0.020	4	3	0.939	1.047	1	1
Marital Status	0.003	0.016	5	4	1.614	1.620	4	3

The rank of the variables are based on VIMP columns for each event

FIGURE 1. Cumulative incidence function of AIDS and death events for HIV data
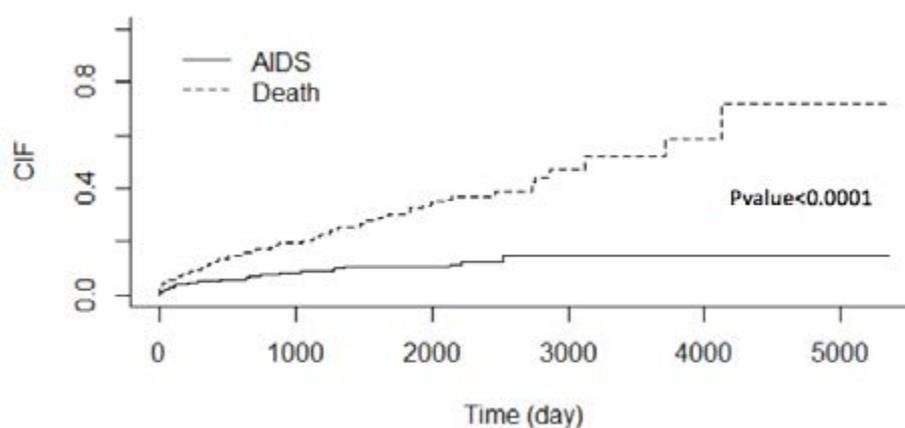
DISCUSSION

Occurring competing events precludes the event of interest in many survival data analysis. In this situation typical survival data analysis cannot be applied. There are several strategies for analyzing survival data for competing risks data analysis. We applied RSF that have been adapted for competing risks data analysis. This method benefits from many useful properties of forests and has several important features. RSF estimates the CIF directly, provides accurate prediction performance, and models non-linear effects as well as interactions. It can be used for event specific selection of risk factors and finally it is free of model assumptions [8, 14].

Results showed that, RSF identified co-infection with TB and gender as the top most important predictors of cause specific hazards for AIDS progression in HIV-infected patients in the presence of competing event of death from causes not related to AIDS. While the two top most important variables directly affected cumulative incidence probability were age at diagnosis and gender respectively. Co-infection with TB variable was also statistically significant cause specific and subdistribution hazards of AIDS progression using traditional competing risks models. However, the second important (significant) variable for both hazards

(cause specific and subdistribution) was transmission type. For the competing event transmission type and gender were two top most important variables for cause-specific RSF and age at diagnosis and transmission type were two top most important variables for subdistribution RSF. Nevertheless, using classical models, age was the only significant variable for cause-specific hazards model and age at diagnosis and transmission type were significant using subdistribution hazards model. These comparisons show a relative consistency between results of the traditional competing risks model strategies and the RSF. We also compared the performance of the methods using cross validation. According to the results, RSF in both cases outperformed classical models in terms of lower prediction error. This can be attributed to the property of considering all complex relationships between variables by the RSF model. Ishwaran et al in their study also showed that their proposed RSF model outperformed traditional models in competing risks case [14].

In the present study, being co-infected with TB was identified as the first top most important variable in AIDS progression which is a leading preventable killer in people who are living with HIV infection. [19]. Our findings indicated that HIV positive subjects who were co-infected with TB had higher risk of incidence of AIDS compared to

TABLE 3. Results of cause specific and subdistribution hazard models for AIDS progression and death competing events

(A) CAUSE-SPECIFIC HAZARDS MODEL						
	AIDS progression			Death from causes not related to AIDS		
Variables	B (SE)	HR	Pvalue	B (SE)	HR	Pvalue
Co-infection (HIV/TB)						
TB negative	Ref.	1	-	Ref.	1	-
TB positive	2.23 (0.035)	9.32	<0.001	-0.81 (0.71)	0.44	0.255
Gender						
Male	Ref.	1	-	Ref.	1	-
Female	-1.06(0.76)	0.34	0.165	-1.39 (0.84)	0.25	0.099
Transmission						
Injecting drug users	Ref.	1	-	Ref.	1	-
Sexual	0.44 (0.59)	1.55	0.459	-0.61 (0.54)	0.54	0.255
Mother to child	2.60 (1.23)	13.43	0.035	-1.30 (1.63e+3)	0.00	0.994
Injecting drug users/Sexual	0.63 (0.61)	1.89	0.295	-9.63 (0.72)	0.38	0.178
Age at diagnosis	0.03(0.02)	1.03	0.155	0.04 (0.11)	1.05	0.001
Marital Status						
Single	Ref.	1	-	Ref.	1	-
Married	0.22 (0.39)	1.25	0.565	0.07 (0.21)	1.07	0.731
Divorced	0.60 (0.43)	1.82	0.156	0.21 (0.28)	1.23	0.448
Widow	0.15 (1.06)	1.16	0.887	-0.07 (0.61)	0.93	0.906
(B) FINE AND GRAY MODEL						
	AIDS progression			Death from causes not related to AIDS		
Variables	B (SE)	HR	Pvalue	B (SE)	HR	Pvalue
Co-infection (HIV/TB)						
TB negative	Ref.	1	-	Ref.	1	-
TB positive	2.30 (0.32)	9.93	<0.001	-1.35 (0.74)	0.26	0.066
Gender						
Male	Ref.	1	-	Ref.	1	-
Female	-1.02 (0.68)	0.36	0.140	-1.38 (0.76)	0.25	0.071
Transmission						
Injecting drug users	Ref.	1	-	Ref.	1	-
Sexual	0.54 (0.53)	1.72	0.300	0.21 (0.49)	0.54	0.210
Mother to child	2.39 (1.08)	10.88	0.027	0.00 (0.63)	0.00	0.001
Injecting drug users/Sexual	0.75 (0.53)	2.11	0.160	0.12 (0.65)	0.37	0.120
Age at diagnosis	0.02 (0.02)	1.02	0.230	0.04 (0.01)	1.04	0.001
Marital Status						
Single	Ref.	1	-	Ref.	1	-
Married	0.20 (0.36)	1.22	0.570	0.102 (0.22)	1.11	0.640
Divorced	0.53 (0.41)	1.71	0.200	0.12 (0.30)	1.14	0.660
Widow	0.12 (1.07)	1.13	0.910	-0.04 (0.61)	0.96	0.940

FIGURE 2. Three-year predicted cumulative incidence function for AIDS progression.
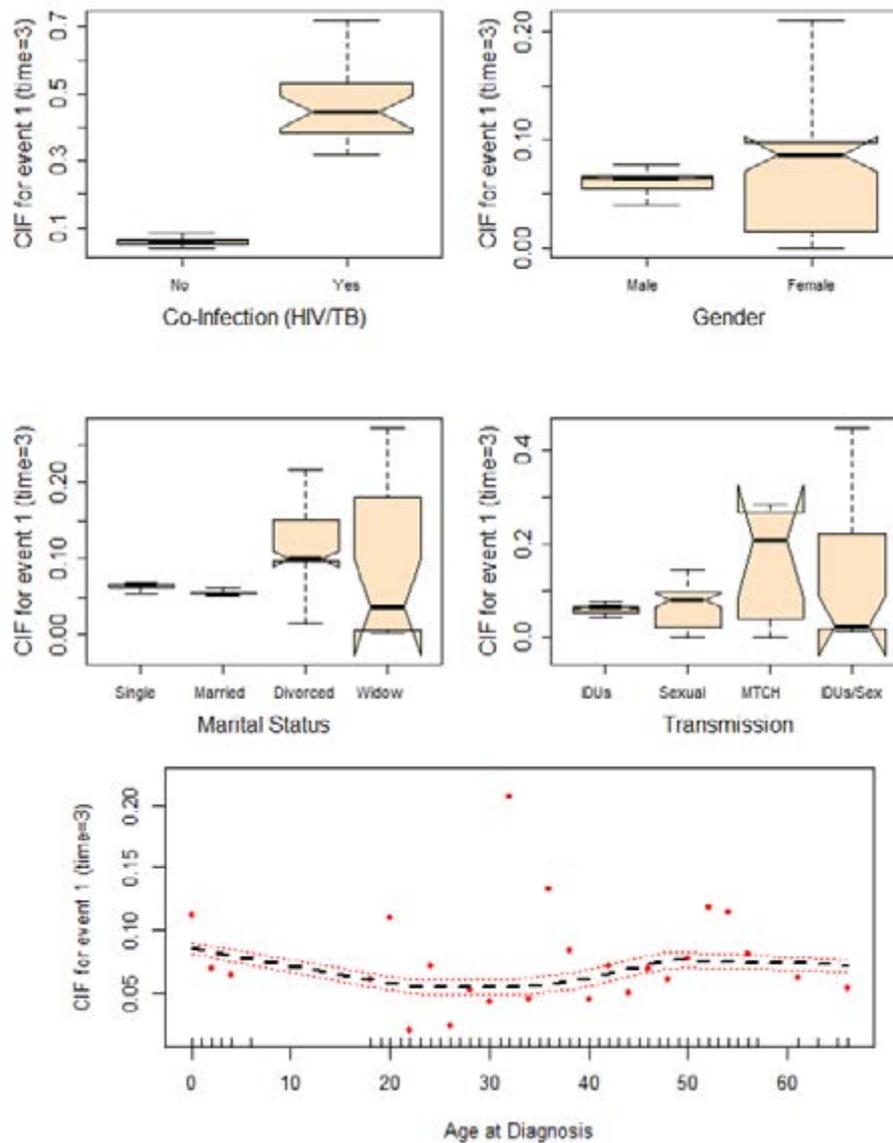
those who infected with HIV alone which is in concordance with the results of other epidemiological studies [5, 6, 20]. Therefore, the importance of treatment of TB in HIV infected people is revealed by this evidence. "In 2004, an interim policy on collaborative TB/HIV activities and emphasized on three distinct objectives was published by the World Health Organization (WHO): (a) establishing and strengthening mechanisms for integrated delivery of TB and HIV services; (b) reducing the burden of TB among people living with HIV and initiating early antiretroviral therapy and (c) reducing the burden of HIV among people with presumptive TB and diagnosed TB" [5].

According to our findings, the predicted cumulative incidence function for progression to AIDS was higher in women compared with men. This finding is in agreement with the result of other studies [5]. The observed difference between the two genders might be explained by two reasons: the small sample size of women (resulting in random errors) and higher proportion of censorship in men compared with women (25% vs. 3%, respectively). Therefore, interpretation of the cumulative incidence between men and women must be done with caution.

According to the results, a large proportion of the HIV infected subjects were male IDUs instead of sexual intercourse. This might be due to the fact that sexual contact with men or non-marital intercourse with women in Islamic Republic of Iran is strongly forbidden for men [5]. There are evidences that show the risk of HIV transmission among IDUs can be successfully reduced by access to sterile injecting equipment, methadone maintenance therapy and outreach services [5]. In addition, the cumulative incidence probability of AIDS progression was

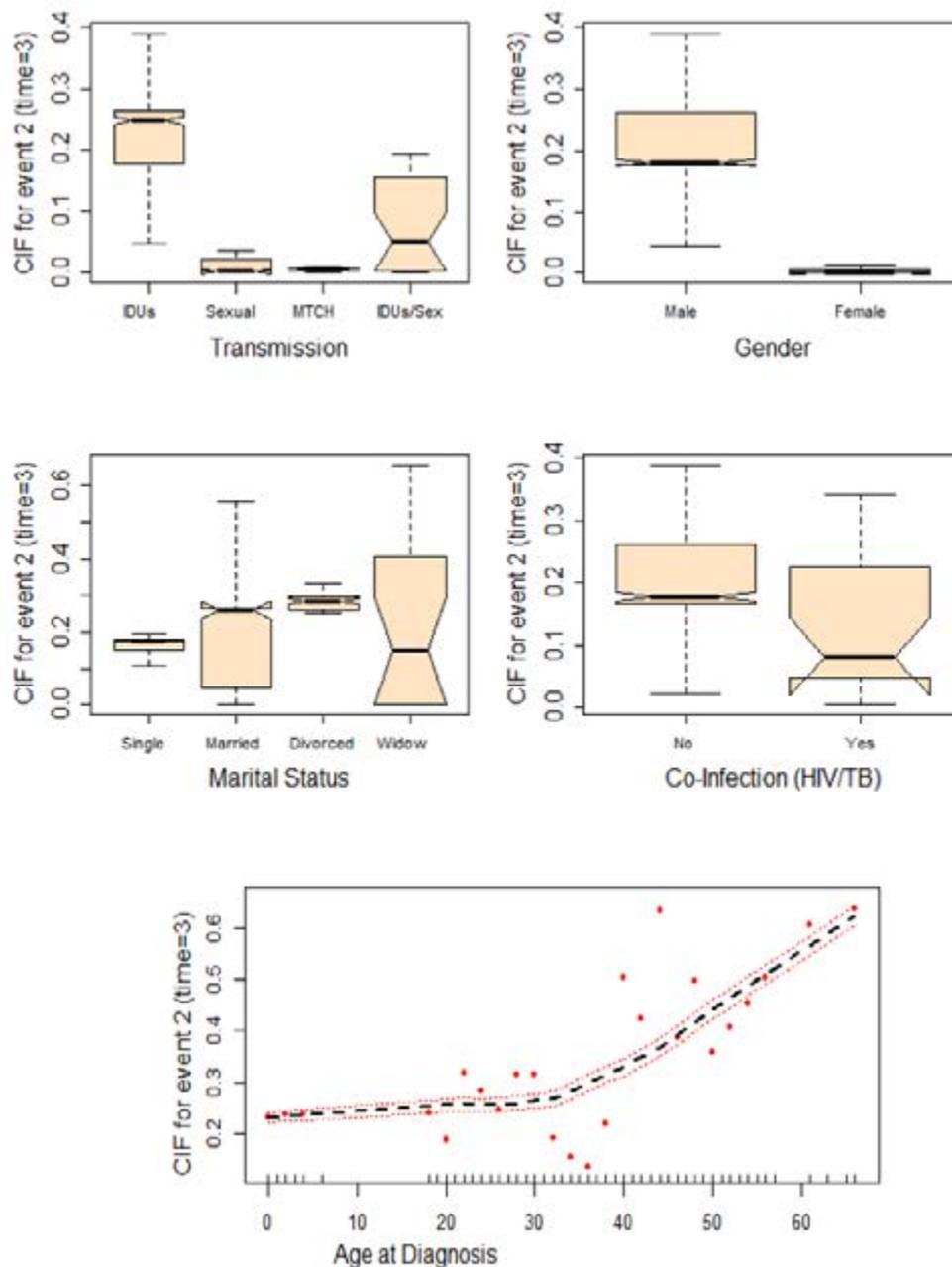**FIGURE 3. Three-year predicted cumulative incidence function for deaths not related to AIDS**

TABLE 4. Comparison between used methods over 100 repetitions using integrated Brier score (IBS)

	IBS	
Variables	**AIDS (Standard deviation)**	**Death(Standard deviation)**
Fine and Gray model	0.082 (0.003)	0.191 (0.004)
RSF1*	0.061 (0.002)	0.152 (0.002)
Cause Specific 1	0.084 (0.008)	0.197 (0.006)
RSF2**	0.063 (0.003)	(0.003)

*and ** are related to RSFs using modified Gray splitting rule and generalized log-rank splitting rule;

FIGURE 4. Prediction error curves for AIDS deaths not related to AIDS using applied methods.
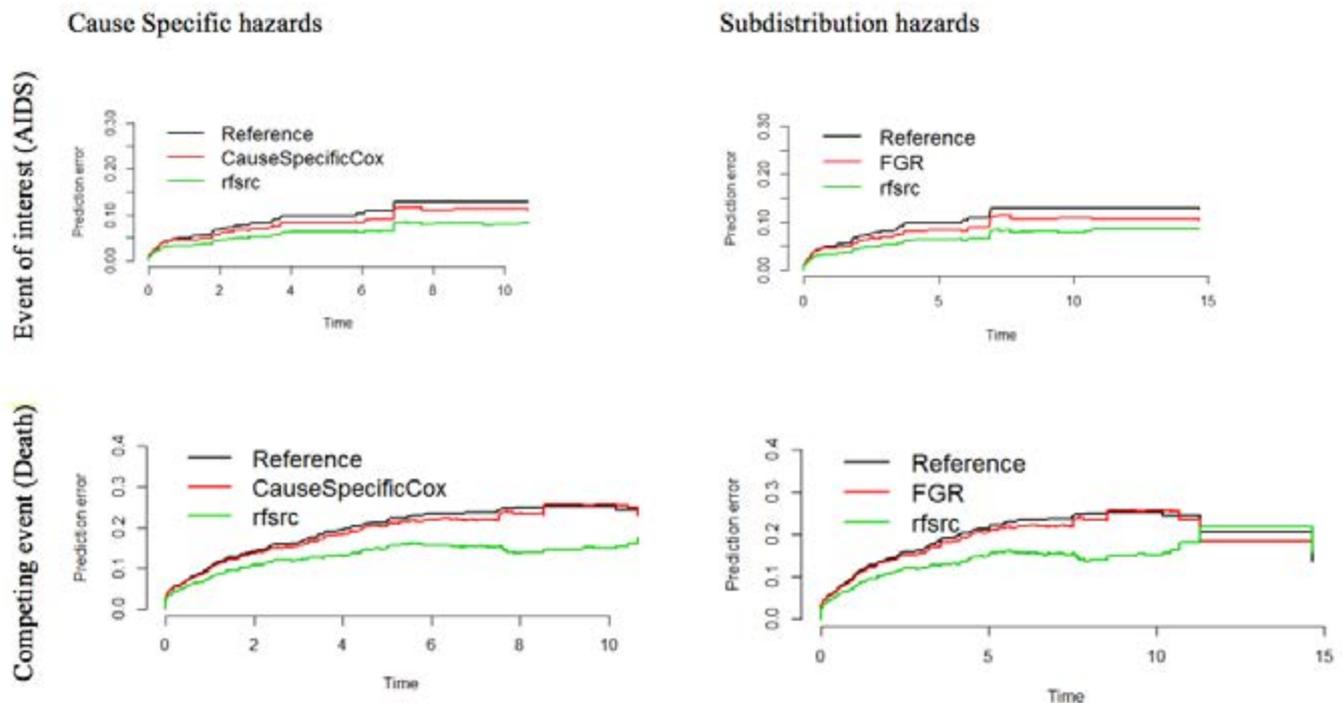
much higher in mother to child transmission compared with other types of transmission which can be attributed to weak immune system of babies. There are several strategies to prevent the risk of mother to child transmission of HIV such as receiving HIV medicines by pregnant women during their pregnancy and using a scheduled cesarean delivery (sometimes called a C-section) in some situations as well as using safe and healthy alternatives of breast milk instead of breastfeed women's babies (because HIV can be transmitted in breast milk) [21].

There were a few limitations in the present study. Survival data analysis including competing risks data requires reliable sources of data based on prospective designs. However, the analysis in the present study was based on a retrospective study recorded by registry centers. Consequently, we were not able to verify the accuracy of the data which can lead to potential information bias. Moreover, to estimate the survival time for progression from HIV to AIDS, the date of diagnosis was used as the beginning of the HIV infection. Nonetheless, some subjects may have become infected with HIV before that time which leads to underestimation of the actual duration of AIDS. The other issue was that since there was not sufficient information about receiving ART, we had to exclude this variable from our analysis because of the nature of RSF for competing risks. As it can reduce the risk of AIDS progression as well as death, it is recommended to record information about taking ART in HIV positive patients progressed or not AIDS.

Despite these limitations, the effect of several predictors

on AIDS progression in a high–middle-income country was apparent which may have a number of implications for healthcare policy and can provide useful information for institution of intervention measures to suppress the progression of HIV to AIDS and death as well as to reduce the risk of death among HIV-positive patients. First, we indicated that an HIV infected subjects who co-infected with TB had higher cumulative incidence probability of progression to AIDS compared to those infected with HIV alone which implies the priority of the diagnosis and treatment of co-infection with TB in HIV positive people. This finding was adjusted for the presence of competing risks using adapted random survival forest for competing risks that is not typically considered in the analysis of this kind of data. Second, a majority of the HIV infected subjects were IDUs which revealed the origin and the mode of HIV transmission in the community. So, it shows the need of special attention by the policy makers who plan preventive programs.

CONCLUSION

The focus of the present study was to identify the important prognostic factors that affect the duration of time from HIV infection to AIDS progression in the presence of a competing event of death from causes not related to AIDS. The results showed that several modifiable and non-modifiable predictors including co-infection with TB and transmission mode affect the risk of progression of AIDS.

Acknowledgement

Conflict of interest

The authors declare no conflict of interest.

References

1. World Health Organization. HIV/AIDS. Retrieved February 2015. 2014a, November 2014; Available from: http://www.who.int/mediacentre/factsheets/fs360/en/.
2. Hamidi O, Tapak L, Poorolajal J, Amini P. Identifying Risk factors for progression to AIDS and mortality post-HIV infection using illness-death multistate model. Clinical Epidemiology and Global Health. 2017.
3. World Health Organization. HIV/AIDS. Retrieved August 4, 2014. (2013, July 2014); Available from: http://www.who.int/mediacentre/factsheets/fs360/en/.
4. Hamidi O, Poorolajal J, Tapak L. Identifying predictors of progression to AIDS and mortality post-HIV infection using parametric multistate model. Epidemiology, Biostatistics and Public Health. 2017.
5. Mirzaei M, Poorolajal J, Khazaei S, Saatchi M. Survival rate of AIDS disease and mortality in HIV-infected patients in Hamadan, Iran: A registry-based retrospective cohort study (1997–2011). International journal of STD & AIDS. 2013:0956462413486457.
6. Poorolajal J, Molaeipoor L, Mohraz M, Mahjub H, Ardekani MT, Mirzapour P, et al. Predictors of progression to AIDS and mortality post-HIV infection: a long-term retrospective cohort study. AIDS care. 2015;27(10):1205-12.
7. Tabarsi P, Chitsaz E, Moradi A, Baghaei P, Farnia P, Marjani M, et al. Treatment outcome, mortality and their predictors among HIV-associated tuberculosis patients. International journal of STD & AIDS. 2012;23(9):e1-e4.
8. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. The annals of applied statistics. 2008:841-60.
9. 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.
10. Chen X, Ishwaran H. Random forests for genomic data analysis. Genomics. 2012;99(6):323-9.
11. Radfar S, Tayeri K, Namdari Tabar H. Practical guidelines on how to provide consulting services in behavioral disorders centers. Tehran: Ministry of Health and Medical Education. 2009.
12. World Health Organization. WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children. 2007.
13. Control CfD, Prevention. Guidelines for national HIV case surveillance, including monitoring for HIV infection and AIDS. Mmwr. 1999;1(48):1-28.
14. Ishwaran H, Gerds TA, Kogalur UB, Moore RD, Gange SJ, Lau BM. Random survival forests for competing risks. Biostatistics. 2014;15(4):757-73.
15. Kleinbaum DG, Klein M. Survival analysis: Springer; 2010.
16. Mogensen UB, Ishwaran H, Gerds TA. Evaluating random forests for survival analysis using prediction error curves. Journal of Statistical Software. 2012;50(11):1.
17. Ishwaran H, Kogalur U. RandomForestSRC: Random forests for survival, regression and classification (RF-SRC). R package version 2.1. 0, 2016.
18. Gray B. cmprsk: Subdistribution analysis of competing risks. URL: http://CRAN R. 2013.
19. Organization WH. WHO policy on collaborative TB/HIV activities. Guidelines for national programmes and other stakeholders. Geneva, Switzerland, 2012. WHO/HTM/TB; 2012.
20. Lopez-Gatell H, Cole SR, Margolick JB, Witt MD, Martinson J, Phair JP, et al. Effect of tuberculosis on the survival of HIV-infected men in a country with low TB incidence. AIDS (London, England). 2008;22(14):1869.
21. Preventing Mother-to-Child Transmission of HIV During Childbirth. Available from: https://aidsinfo.nih.gov/education-materials/fact-sheets/24/70/preventing-mother-to-child-transmission-of-hiv-during-childbirth.

*