Evaluation of Survival Analysis Models for Predicting Factors Infuencing the Time of Brucellosis Diagnosis

Sadegh kargarian-Marvasti, M.Sc. [1], Sima Afrashteh, M.Sc. [2], Gholamreza Rafiei, B.Sc. [3]

1- MPH Student in Epidemiology, Isfahan University of Medical Sciences, Isfahan, Iran

2- PhD student in Epidemiology, Department of Public Health , Faculty of Health ,Bushehr University of Medical Sciences, Bushehr, Iran
(Corresponding author; E-mail: sima.afrashte3@gmail.com)

3- Bsc of public health, Isfahan University of Medical Sciences, Fereydunshahr health center, Isfahan, Iran

ARTICLE INFO

Abstract

Background: Brucellosis or Malta fever is one of the most common zoonotic diseases in the world. In addition to causing human suffering and dire economic impact on animals, due to the high prevalence of Brucellosis in the western regions of Isfahan province, this study aimed to analyze effective factors in the time of Brucellosis diagnosis using parametric and semi-parametric models and to evaluate the goodness of fit of these models.

Methods: This historical cohort study, 412 patients with Brucellosis in Fereydunshahr, Iran who had referred to hospital, rural & urban health centers and physicians' private clinics in Fereydunshahr between 2006 and 2016 were recruited through census sampling. The failure (or event) in this study, was diagnosis of Brucellosis based on positive immunologic tests (2-ME test ≥1:40 and Wright serology ≥1:80). In order to eliminate confounding variables, effective factors of the time of Brucellosis diagnosis were determined using univariate (P≤0.20) and multivariable (P<0.05) analysis according to Cox semi-parametric model and five parametric models (weibull, exponential, log-logic, log-normal and gompertz) and the best fitted model was identified. Data were analyzed using R software version 3.2.3.

Results: According to the results of this study, occupation (farmer and livestock breeder), place of residence (urban), having a history of direct contact with livestock, simultaneous infection in other family members, and the newness of the disease (vs. recurrence) were identified as predictors of early detection of the disease.

Conclusion: Despite the researchers' tendency to use Cox method in survival analysis, in this study, according to AIC, "Gopmpertz" parametric model was recognized as the best fitted regression model in the analysis of the effective factors in the definitive time of Brucellosis diagnosis.

Introduction

Brucellosis or Malta fever is one of the most common zoonotic diseases in the world, which is found in many parts of the world, including Latin America, Middle East, Africa and Asia (1). Annually, more than 500,000 new human cases of brucellosis are reported in the world (2).

One of the most important advantages of survival methods in clinical sciences is their ability to

manage such censored observations which are ignored by other methods (such as logistic or linear regression). Survival analysis is one of the statistical methods that was applied to study for the occurrence and time of occurrence of an event such as death, cancer survival, relapse, etc. (3, 4). There are 2 types of regression models for survival analysis; parametric models (like Exponential, Log-logistic, Weibull, Log-normal and Gompertz models) and cox semi-parametric proportional hazard model (5). Despite some limitations, Cox model is a common method in survival modeling, but parametric models have better efficiency (4, 6), under certain circumstances (6-10). The results of parametric models and cox regression were consistent in Rajaeefard *et al.* (11). In a survival analysis of patients with gastrointestinal cancer, log-logistic parametric model declared as the best-fitted model (12). Viswanathan *et al.* reported cox model as the best model in the study of risk factors of diabetic nephropathy (13). In medical sciences, semi-parametric survival analysis (such as cox model) is usually favored because, usually, data do not gratify the prerequisites of parametric survival analysis. Since models can produce different results depending on the extent of matching between the underlying assumptions of each model and the specific characteristics of the clinical-healthcare problem (14), we compared parametric and semi-parametric models.

Due to human suffering, and economic impact of Brucellosis in the animals, as well as the high prevalence of this disease in the western regions of Isfahan province, this study aimed to analyze effective factors in the time of Brucellosis diagnosis using parametric and semi-

parametric models and to evaluate the goodness of fit of these models.

Materials and Methods

In this historical cohort study, 412 patients with Brucellosis who had referred to Fereydunshahr hospital and its rural & urban health centers and physicians' clinics, between 2006 and 2016, were recruited through census method. They were followed for the diagnosis of Brucellosis. Data were extracted from patients' health-care records and have been presented in table 1. The failure (or event) in this study, according to the national and global protocol, was diagnosis of Brucellosis in patients based on positive immunologic tests (2-ME test ≥1:40 and Wright serology ≥1:80) and clinical symptoms (15-17). Studies have shown that after a 30-day delay in the diagnosis of the disease (after the onset of symptoms), the probability of occurrence of complications due to Brucellosis, increases (18-20). For this reason, in evaluating fitness of survival regression models, patients who had a definite diagnosis of the disease in less than 30 days after the first clinical symptoms of brucellosis were compared with those patients in whom, this period time was more than 30-days. Serum agglutination tests (based on 2-ME and Wright) were measured on blood samples and antibody titers were determined using an antigen kit from the Pasteur Institute (Pasteur Institute of IRAN, Tehran, IR Iran).

In this study, to eliminate confounding variables and to identify the best fitted model, effective factors in the time of Brucellosis diagnosis were determined using

univariate and multivariable analysis according to Cox semi-parametric model and five parametric models (weibull, exponential, log-logic, log-normal and gompertz). Researchers often prefer cox model to the parametric models due to its less assumptions. Although some studies showed that parametric models estimate the variables more efficiently than cox model (4). Since models can produce different results depending on the extent of matching between the underlying assumptions of each model and the specific characteristics of the clinical-healthcare problem (14), we compared these 5 models. After univariate analysis, variables with statistical significance (P≤0.20) were entered in a multiple regression model (P<0.05) and analyzed using Back-ward: LR stepwise method.

Risk Ratio (RR) was calculated in Cox and Gompertz models as Hazard Ratio (HR), as well as, in weibull, exponential, log-logic and log-normal models as Time Ratio (TR) (Tables 2 & 3). HR and RR have similar interpretation, but HR gives instantaneous risk at a particular time and RR gives cumulative risk over a time span. HR is the probability of an event at a particular time, provided that intended event has not occurred, before that time. HR>1 shows that the group is more high risk for the occurrence of the event versus the reference group.

In this study, Kaplan-Meier's nonparametric approach was used to compare two survival functions using Mantel-Cox Log Rank Test. This test compares the number of observed final outcomes in each group with the number of expected final outcomes (similar to Chi-square test and rejection of the zero hypothesis with P<0.05) (21).

Also, we used Akaike's information criterion AIC (Akaike's Information Criterion) and standardized variation of parameters to evaluate the goodness of fit (22). The goodness of fit of a statistical model describes how well it fits a set of observations. This criterion was formulated by the statistician "Hirotugu Akaike" (1974) in the following equation; given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Thus, AIC rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters (4):

$$AIC = -2 \times \log(\text{likelihood}) + 2 \times (a + c)$$

In this formula, "a" is the number of parameters, and "c" is a constant factor (for example, equals 0 in the cox model, equals 1 in the exponential model and equals 2 in the Weibull, log-logistic and log-normal models) (22).

Results

A total of 412 patients entered the study, of whom 148 (35.9%) were female and 264 (64.1%) were male (Table 1). Mean age of subjects was 30.41±0.91 years (ranged from 1 to 89 years and median age of 25 years). Non-parametric Kaplan-Meier approach showed that the mean of diagnosis time of brucellosis was 19.00±1.30 days after the first clinical symptoms [95% CI: 16.46-21.54]. According to the results, the time of diagnosis has been less than 30-days in 62.4% of patients (N=257) and the rest of patients were considered as censored observations.

Table 1. Frequency distribution of demographic variables and risk factors of patients with brucellosis

Variable		n	%	Variable		n	%
Age	<10 years	37	9.0	Disease status	New	382	92.7
	10-19 y	97	23.5		Relapse	30	7.3
	20-29 y	105	25.5	Contact with animals	Yes	366	88.8
	30-39 y	61	14.8		No	46	11.2
	40-49 y	37	9.0	Type of unpasteurized dairy products	Milk	216	52.4
	50-59 y	33	8.0		Cheese	18	4.4
	≥60 years	42	10.2		Milk + Cheese	90	21.9
Gender	Male	264	64.1		Other	3	0.7
	Female	148	35.9		Not used	85	20.6
Job	Farmer	57	13.8	Time between onset of symptoms and diagnosis	<1 month	257	62.4
	Stockbreeder	107	26.0		≥1 month	155	37.6
	Housewife	120	29.1	Livestock vaccination	Yes	306	74.3
	Student	58	14.1		No	71	17.2
	Child	26	6.3		Without livestock	35	8.5
	Other	44	10.7	Season of the event of disease	Spring	133	32.3
Education	Illiterate	173	42.0		Summer	190	46.1
	Elementary	105	25.5		Fall	51	12.4
	middle school	95	23.0		winter	38	9.2
	Diploma or high	39	9.5	Keeping livestock at home	Yes	376	91.3
Habitat	Urban	74	18.0		No	36	8.7
	Rural	201	48.8	Ethnicity	Lor	305	74.0
	Tribal	137	33.2		Turk	69	16.8
Infection of family members	Yes	163	39.6		Georgian	31	7.5
	No	249	60.4		Fars	7	1.7

Table 2 shows the results of multivariable analysis of parametric and semi-parametric models for significant variables. Table 3 shows the result of univariate analysis.

We used Kaplan-Meier's nonparametric approach to compare two survival functions. Figures 1 & 2 show comparison of two significant variables (type of disease and direct contact with animal) based on non-parametric Kaplan-meier method. For example Kaplan–Meier survival analysis showed that the event time of brucellosis was shorter in new cases compared to recurrence cases. This time was shorter in patients who had direct contact with animal than in patients without contact (figure 1 & 2).

Table 2. Comparison of the final results of cox and parametric models in multivariable analysis for the diagnosis of brucellosis (P<0.05)

Variables	Log-Normal Reg.			Weibull Reg.			Log-Logistic Reg.			Gompertz Reg.			Exponential Reg.			Cox Regression		
	TR	95% CI	P	HR	95% CIyy	P	TR	95% CI	P	HR	95% CI	P	HR	95% CI	P	HR	95% CI	P
Occupation	No significance			1.13	1.04 1.22	0.003	No significance			1.09	1.00 1.17	0.04	1.15	1.06 1.24	0.0001	No significance		
place of residence	No significance			No significance			No significance			No significance			No significance			1.88	1.03 3.44	0.40
Contact with livestock	1.87	1.05 3.35	0.03	1.71	1.11 2.65	0.02	1.90	1.05 3.45	0.03	1.56	1.01 2.42	0.04	No significance			1.58	1.01 2.47	0.04
Infection in family	1.63	1.14 2.34	0.008	1.48	1.14 1.94	0.004	1.56	1.08 2.23	0.02	1.37	1.05 1.79	0.02	1.59	1.22 2.07	0.001	1.30	1.00 1.69	0.05
Type of the disease	6.66	3.17 14.0	0.0001	5.23	2.65 10.3	0.0001	8.41	3.55 19.9	0.0001	2.83	1.45 5.53	0.002	9.02	4.60 17.7	0.0001	3.00	1.54 5.86	0.001

Table 3. Comparison of the final results of cox and parametric models in univariable analysis for diagnosis of brucellosis (P≤0.20)

	Log-Normal Reg.			Weibull Reg.			Log-Logistic Reg.			Gompertz Reg.			Exponential Reg.			Cox Regression		
	TR	95% CI	P	TR	95% CI	P	TR	95% CI	P	HR	95% CI	P	TR	95% CI	P	HR	95% CI	P
Age	No Significance															1.70	1.04 2.79	0.20
Sex	1.40	0.96 2.06	0.08	1.73	1.12 2.69	0.01	1.46	1.00 2.15	0.05	1.29	0.99 1.68	0.06	1.57	1.21 2.05	0.0001	1.25	0.96 1.63	0.09
Occupation	1.12	1.00 1.25	0.05	1.18	1.05 1.34	0.01	1.12	1.01 1.25	0.04	1.08	1.00 1.16	0.05	1.18	1.09 1.27	0.0001	1.61	1.06 2.45	0.20
Habitat	No Significance															1.23	0.87 1.73	0.18
Ethnicity	1.31	1.01 1.69	0.04	1.44	1.08 1.91	0.01	1.30	1.02 1.67	0.04	1.17	0.99 1.39	0.06	1.42	1.19 1.69	0.0001	1.49	0.28 1.63	0.16
Contact with livestock	1.47	0.82 2.65	0.20	1.63	0.81 3.29	0.17	1.55	0.85 2.83	0.16	1.32	0.87 2.02	0.19	No Significance			1.33	0.87 2.02	0.18
consuming unpasteurized dairies	1.41	0.90 2.19	0.13	1.70	1.05 2.77	0.03	1.39	0.90 2.14	0.13	1.24	0.93 1.66	0.15	1.76	1.31 2.36	0.0001	1.25	0.94 1.68	0.13
Infection in family	1.75	1.21 2.54	0.003	2.05	1.34 3.15	0.001	1.75	1.20 2.54	0.003	1.42	1.10 1.84	0.01	1.76	1.35 2.28	0.0001	1.39	1.07 1.79	0.01
Type of the disease	7.52	3.55 15.9	0.0001	14.7	5.29 40.7	0.0001	10.2	4.19 24.8	0.0001	3.09	1.59 6.03	0.001	10.3	5.29 20.0	0.0001	3.07	1.58 5.98	0.001

Survival function based on type of disease

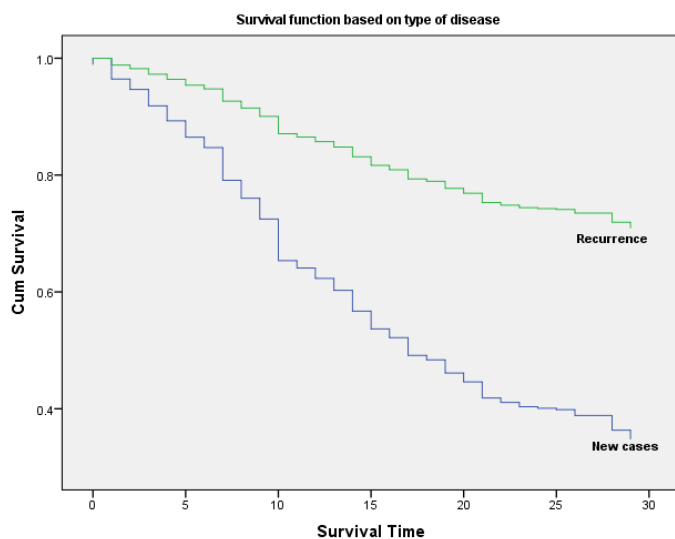**Figure 1.** Cumulative survival function in patients with brucellosis in accordance with the type of disease

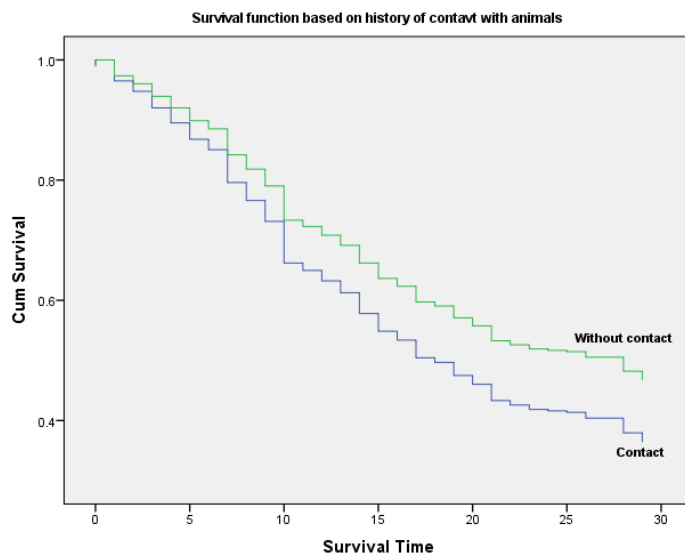**Survival function based on history of contavt with animals**

Figure 2. Cumulative survival function in patients with brucellosis in accordance with direct contact with livestock

Although univariate analysis results were not different between parametric and semi-parametric models, based on AIC, parametric models showed a preferable fit to our data than semi-parametric Cox model (Table 4). The "Gompertz parametric model" with the lowest AIC's value, provided the excellent goodness of fit to the data.

Table 4. Comparison of fitness of models based on AIC

Model	-2 × Log Likelihood	AIC
Cox	2895.45	2903.45
Exponential	1383.70	1391.70
Weibull	1296.13	1308.13
Log-logistic	1229.67	1239.67
Log-normal	1224.18	1234.19
Gompertz	1145.37	1153.37

AIC: Akaike information criterion

Discussion

The main purposes of this study were investigating effective factors in the definitive time of Brucellosis diagnosis using semi-parametric and parametric models, and, compare the fitness of these models based on AIC.

According to the results of this study, direct contact with livestock was one of the effective factors on the detection of brucellosis. Other studies have pointed Contact with livestock as an important reservoir for the disease (23, 24). One study in Iran showed that exposure to animals increases the chance of developing this disease (23). Cash-Goldwasser et al. showed association between livestock and brucellosis (24). Our findings showed simultaneous presence of infection in the other family members as an effective factor in determining time of brucellosis diagnosis. El-Koumi et al. showed that 45% of children with brucellosis had a positive family history of the disease (25). Contrary to our study, in some studies, the family history of the disease has not been associated with the disease (23).

In this study, time of diagnosis in 62.4% of patients (N=257) was less than 30-days. According on non-parametric Kaplan-Meier approach, the median time of brucellosis diagnosis was 19.00±1.30 days after the first

clinical symptoms [95% CI: 16.46-21.54]. Also, "Gompertz model" provided the excellent goodness of fit to our data. In order to minimize selection bias and external validity increase, we used census method in this study.

A number of studies have been directed to compare several survival models, which some suggested semi-parametric models as the most appropriate modeling method (13, 26), and some implied parametric models (4, 27).

In this study, parametric models had a better fitness than Cox model. Several studies parallel with this study showed that fitness of parametric methods were better than cox regression. Kargarian et al. (4) evaluated log-normal model in survival analysis of event time of neuropathy in patients with type 2 diabetes. Roshany et al. assessed Weibull model as the best-fitted model in the application of parametric, semi-parametric and nonparametric approaches in survival analysis of patients with acute myocardial infraction (27). However, in some studies, the results of data analysis were approximately similar in both Cox model and parametric models(11). In contrast with our study, some studies such as a study on

diabetes-related lower-extremity amputation by Lacle *et al.* (26) and a study about risk factors associated with the development of overt nephropathy in type 2 diabetes patients by Viswanathan *et al.* (13) showed a better fitness for Cox model than parametric models. We did not find any study about survival analysis of Brucellosis to compare with our study.

Based on the results of this study, in case of existing a patient in a family, education about the transmission ways of the disease to other members of family, and more importantly, referring the patient with clinical symptoms to a specialist are highly recommended. Since women, especially in rural areas, are the main pillar of the family in the process of keeping animals, milking and preparing dairy products, we emphasize on the training of those housewives who have contact with animals.

Conclusion

Despite the researchers' tendency to use Cox method in survival analysis, parametric models have more precise results than Cox model, especially, when fewer censored data are presented. In this study, according to AIC, "Gopmpertz" parametric model, was recognized as the best fitted regression model in the analysis of the effective factors in the definitive diagnosis time of Brucellosis.

Financial support and sponsorship

Nil

Conflicts of interest

The authors have no conflicts of interest.

References

1. Adesokan HK, Alabi PI, Ogundipe MA. Prevalence and predictors of risk factors for Brucellosis transmission by meat handlers and traditional healers' risk practices in Ibadan, Nigeria. J Prev Med Hyg 2016; 57(3):E164-71.

2. Golshani M, Buozari S. A review of brucellosis in Iran: epidemiology, risk factors, diagnosis, control, and prevention. Iran Biomed J 2017; 21(6):349-59.

3. Kargarian Marvasti S, Abolghasemi J, Heydari I, Rimaz S. Effective factors in the time of development of neuropathy in type ii diabetic patients. Iranian Journal of Epidemiology 2017; 13(2):80-9. [In Persian].

4. Kargarian-Marvasti S, Rimaz S, Abolghasemi J, Heydari I. Comparing of cox model and parametric models in analysis of effective factors on event time of neuropathy in patients with type 2 diabetes. J Res Med Sci 2017; 22:115.

5. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. New York: Springer; 2000.

6. Efron B .The efficiency of Cox's likelihood function for censored data. J Am Stat Assoc 1977; 72(359):557-65.

7. Lawless JF. Parametric Models in Survival Analysis. In: Armitage P, Colton T. Encyclopedia of Biostatistics. 2nd ed. New York: Wiley; 2005.

8. Bennett S. Analysis of survival data by the proportional odds model. Statistics in Medicine 1983; 2(2):273-77.

9. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. 2nd ed. Wiley: New York; 2002.

10. Oakes D. The asymptotic information in censored survival data. Biometrika 1977; 64(3):441-8.

11. Rajaeefard A, Moghimi Dehkordi B, Tabatabaee HR, Zeighami B, Safaee A, et al. Applying parametric models for survival analysis of gastric cancer. FEYZ 2009; 13(2):83-8. [In Persian].

12. Ghadimi M, Mahmoodi M, Mohammad K, Zeraati H, Rasouli M, Sheikhfathollahi M. Family history of the cancer on the survival of the patients with gastrointestinal cancer in northern Iran, using frailty models. BMC Gastroenterol 2011; 11:104.

13. Viswanathan V, Tilak P, Kumpatla S. Risk factors associated with the development of overt nephropathy in type 2 diabetes patients: a 12 years observational study. Indian J Med Res 2012; 136(1):46-53.

14. Gregori D, Petrinco M, Bo S, Desideri A, Merletti F, Pagano E. Regression models for analyzing costs and their determinants in health care: an introductory review. Int J Qual Health Care 2011; 23(3):331-41.

15. Gomez MC, Nieto JA, Rosa C, Geijo P, Escribano MA, Munoz A, et al. Evaluation of seven tests for diagnosis of human brucellosis in an area where the disease is endemic. Clin Vaccine Immunol 2008; 15(6):1031-3.

16. Gavazzi G, Prigent D, Baudet JM, Banoita S, Daoud W. Epidemiologic aspects of 42 cases of human brucellosis in the Republic of Djibouti. Med Trop (Mars) 1997; 57(4):365-8.

17. Honarvar B, Moghadami M, Lankarani KB, Davarpanah MA, Ataolahi M, Farbod A, et al. Brucellosis as a neglected disease in a neglected population: a seroepidemiological study of migratory nomads in the Fars province of Iran. Epidemiol Infect 2017; 145(3):491-7.

18. Gooya MM, Ahmadnia H, Ghotbi M, Rahimi F, Pezeshki Z. A Comprehensive Guide to the Infectious Disease Care System for Family Physicians. Ministry of Health, Treatment and Medical Training Communicable Disease Management Center 2012; 1:153-62.

19. Colmenero JD, Reguera JM, Martos F, Sanchez-De-Mora D, Delgado M, Causse M, et al. Complications associated with Brucella melitensis infection: a study of 530 cases. Medicine (Baltimore) 1996; 75(4):195-211.

20. Eales KM, Norton RE, Ketheesan N. Brucellosis in northern Australia. Am J Trop Med Hyg 2010; 83(4):876-8.

21. Ahmadi A, Mobasheri M, Hashemi-Nazari SS, Baradaran A, Choobini ZM. Prevalence of hypertension and type 2 diabetes mellitus in patients with colorectal cancer and their median survival time: a cohort study. J Res Med Sci 2014; 19(9):850-4.

22. Klein JP, Moeschberger LM. Survival Analysis: Techniques for Censored and Truncated Data. 2nd ed. New York: Springer. 2005.

23. Alavi SM, Mugahi S, Nashibi R, Gharkholu S. Brucellosis risk factors in the southwestern province of Khuzestan, Iran. Int J Entric Pathog 2014; 2(1):e15610.

24. Cash-Goldwasser S, Maze MJ, Rubach MP, Biggs HM, Stoddard RA, Sharples KJ, et al. Risk factors for human brucellosis in Northern Tanzania. Am J Trop Med Hyg 2017; 98(2):598-606.

25. El-Koumi MA, Afify M, Al-Zahrani SH. A prospective study of brucellosis in children: relative frequency of pancytopenia. Mediterr J Hematol Infect Dis 2013; 5(1):e2013011.

26. Lacle A, Valero-Juan LF. Diabetes-related lower-extremity amputation incidence and risk factors: a prospective seven-year study in Costa Rica. Rev Panam Salud Publica 2012; 32(3):192-8.

27. Roshany D, Azadi NA, Esmail-Nasab N, Yaghoubi M. Application of parametric, semiparametric and nonparametric approaches in survival analysis of patients with acute myocardial infarction. Journal of North Khorasan University of Medical Sciences 2012; 3(5):45-51. [In Persian].