

Cox Proportional Hazard Regression for Risk Factors of Alzheimer's Disease

Nargess Hosseinioun¹

Abstract:

Background: Alzheimer's disease is a form of dementia, mainly strikes people in their 60 and 70s. While no one cause has been determined, researchers have identified certain factors which may put people at a higher risk of developing the disease. Iran's Alzheimer Association has announced that more than 35% of people over 80 years old suffers from this problem across the country. This paper aims to investigate modifiable risk factors of Alzheimer's disease based on Kaplan-Meier estimator and Cox regression analysis (proportional hazard model) and to find out which factors are related to developing of this disease.

Materials & methods: Residents newly admitted to nursing homes with a diagnosis of probable Alzheimer's disease have been considered. Patients were at least 75 years of age from 9 Medicare/Medicaid certified nursing homes of Khorasan and Tehran states and we excluded patients with a history of mental retardation, mental illness or any other long-life mental health disorders. This strategy yielded a sample of 115 patients (37 Males and 78 Females) with total death rate of 47%.

Results: Our findings demonstrate that Age, Gender, Heritage, Apoplexy, Mental and Physical Activities statistically affect Survival and also Hazard Rate, In contrast, we found no link between survival duration and a history of Addiction and Blood Pressure.

Conclusions: These findings have implications for clinical practice and intervention strategies in medical and public health.

Keywords: Survival Analyses, Brain disorder, Alzheimer, Cox proportional hazard regression.



DOI: <https://doi.org/10.3329/jom.v20i2.42006>

Copyright: © 2019 Hosseinioun N. This is an open access article published under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not changed in any way and it is not used for commercial purposes.

Received: 23 May, 2018;

Accepted: 12 September, 2018

Introduction:

The number of Iranians living with Alzheimer's disease (AD) is growing fast. It affects one in 40 Iranian Families. Iran, which had faced young population growth, is experiencing elderly population growth in the last two decades. Iran's Alzheimer Association has announced that more than 35% of people over 80 years old suffer from this problem across the country. AD is a kind of brain disorder which leads to a slow decline in thinking and reasoning skills. In this disease damage and death of brain cells lead to dysfunction of memory, thought, judgment, language, and other nervous activities which consequently change individual behavior and personality.^{1,2,3}

The displeasing outcomes lead to higher healthcare expenses and bring emotional challenges for patients' sibling and

relatives with subsequent mental as well as spiritual burden for the whole families. These outcomes are so wide and painful from time to time that can affect active and productive sections of the society seriously and cause in high cost on the communities.

Various demographic, environmental and clinical risk factors have been diagnosed as plausible factors of Alzheimer's disease (AD). Examples include lack of exercise, obesity, smoking or exposure to secondhand smoke, high blood pressure, high blood cholesterol, poorly controlled type 2 diabetes, and diet lacking in fruits and vegetables. With the exception of cases of Alzheimer's caused by genetics, studies have revealed that Alzheimer's, like other common chronic diseases, develops as a result of multiple factors rather than a single one. Moreover some studies have shown a link between Alzheimer's disease and a major head injury and Apoplexy.^{4,5}

Numerous studies are available concerning various risk factors are related to AD such as obesity,⁶ smoking, diabetes,

1. Department of Statistics, Payame Noor University, Iran.

Corresponding author: Nargess Hosseinioun, Department of Statistics, PayameNoor University, 19395-4697, Iran. Email: Narges_8h@yahoo.com.

hypertension (HTN) in midlife.^{7,8} It is also declared that physical activity would be a protective factor against AD.⁷⁻¹⁰

In this research we aim to identify risk factors associated with Alzheimer's disease, and to evaluate whether these factors vary according to severity of cognitive impairment. Data were from the database which includes information on 115 residents admitted between 2010 and 2015 to 9 Medicare/Medicaid certified nursing homes of Khorasan⁴ and Tehran (Iran) states. Patient data including demographic characteristics like Age, Gender, Addiction, Blood Pressure, physical and mental activities and other clinical and treatment variables were collected with the Minimum Data Set.

Material and methods:

Statistical Analysis

Let T represent survival time as a random variable with cumulative distribution $P(T) = \Pr(T \leq t)$ and probability density function $p_t = \frac{dP_t}{dt}$. Then the survival function S(T) is the complement of the distribution function ($S(T) = \Pr(T > t) = 1 - P(t)$). The hazard function, which assesses the instantaneous risk of demise at time t, conditional on survival to that time is defined as:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t)}{\Delta t} = \frac{f(t)}{S(t)}$$

Models for survival data usually utilize the hazard function or the log hazard. A common hazard models includes

$$\log h(t) = \vartheta + \rho t$$

which reaches the Gompertz distribution of survival times, and

$$\log h(t) = \vartheta + \rho \log t$$

which is Weibull distribution of survival times. In both above distributions, the hazard can either increase or decrease with time; moreover, in both instances, one may find the exponential model when $\rho = 0$.

Censoring is said to be present when information on time to outcome event is not available for all study participants. A datum T_i is said to be right-censored if the event occurs at a time after a right bound, but we are not sure when. The only information we have is this right bound. This is very important in study of survival time, because data are often right-censored. On the contrary a datum T_i is called left-censored if we know that the event occurs at a time before a left bound, but we don't know when. It happens, for example, when the date of a medical exam that revealed a disease is available, but we don't know when the patient has been infected and finally if the event occurs in a time interval (L_i, R_i) , but we don't know exactly when in this interval then the datum T_i is said to be interval-censored.

Now the connection of the survival distribution to covariates is examined by Survival analysis. This examination basically entails the specification of a linear-like model for the log hazard. For example, a parametric model based on the exponential distribution may be given by

$$h_i(t) = \exp(\alpha(t) + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}),$$

that is, as a linear model for the log-hazard or as a multiplicative model for the hazard. Here, i is a subscript for observation, and the x_i 's are the covariates. Then the Cox model is named the *hazard function* denoted by $h(t)$, leaves the baseline hazard function $\alpha(t) = \log h_0(t)$, can be estimated as follow:

$$h_i(t) = h_0(t) \exp(\beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})$$

This model is semi-parametric because while the baseline hazard can take any form, the covariates enter the model linearly. Consider, now, two observations and that differ in their x-values, with the corresponding linear predictors

$$\gamma_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$$

And

$$\gamma_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$$

Then the ration of hazard function for these two observations is independent of time t and given by

$$\frac{h_i(t)}{h_i(t)} = \frac{h_0(t)e^{\gamma_i}}{h_0(t)e^{\gamma_i}}$$

The ratio of hazard functions can be regarded as a ratio of risk functions, so the proportional hazards regression model can be applied as a function of relative risk (while logistic regression models are a function of an odds ratio). Interpreting the Cox model involves examining the coefficients for each explanatory variable.^{11,12}

Cox regression model is method for examining the effect of several variables upon the time a specified event takes to happen. In the context of an outcome such as death this is known as Cox regression for survival analysis. The method does not include any particular "survival model" but it is not truly nonparametric because it does assume that the effects of the predictor variables upon survival are constant over time and are additive in one scale. A positive regression coefficient for an explanatory variable leads to the fact that the hazard is higher and thus the prognosis worse. Conversely, a negative regression coefficient implies a better prognosis for patients with higher values of that variable. When it is used to analyses

the survival of patients in a clinical trial, the model grants us to isolate the effects of treatment from the effects of other variables.^{10,13}

Data Collection

This study was a national survey, carried out in 2 main provinces of Iran, including Tehran and Khorasan in 2015. Residents newly admitted to nursing homes with a diagnosis of probable Alzheimer’s disease have been selected. The key criteria of the diagnostic and statistical manual of mental disorders (DSM-IV) were used for the definition of Alzheimer’s disease. Patients were at least 75 years of age from 9 Medicare/ Medicaid certified nursing homes and we excluded patients with a history of mental retardation, mental illness or any other long-life mental health disorders. Over 150 participants were enrolled in the study; a total of 35 individuals were excluded from the analysis due to incomplete information. Mean age of the participants was 81.22 years while Median was 63.00 years old and standard deviation was 7.227. The data was collected through interviews with the participants or one of their informant family members. To collect demographic information such physical and mental activities, we used an expert approved forms. Information on death was derived through linkage to Medicare files. The study population response rate was 91%. All these strategy yielded a sample of 115 patients (37 Males and 78 Females) with the following information:

Table I: Case Processing Summary

Gender	N	Death	Censoring	Mean	Median
Male	37	17	20(54.1%)	64.066	53.000
Female	78	44	34(43.6%)	83.674	63.000
Total	115	61	54(47%)	81.222	63.000

Results:

Residents newly admitted to nursing homes with a diagnosis of probable Alzheimer’s disease have been selected. The key criteria of the diagnostic and statistical manual of mental disorders (DSM-IV) were used for the definition of Alzheimer’s disease. Patients were at least 75 years of age from 9 Medicare/ Medicaid certified nursing homes of Khorasan and Tehran states, and we excluded patients with a history of mental retardation, mental illness or any other long-life mental health disorders. The significant level were set as $\alpha < 0.05$.

Table II: Survival Information

Factors	Exp(B)	Log Rank Test(sig)	Sig(B)
Gender	1.172	.020	.000
0:Male			
1:Female			
Age			
<75			
75-89	1.190	.001	.001
>89			
Heredity			
0:with	.451	.000	.012
1:without			
Addiction			
0:with	.835	.540	.562
1:without			
Apoplexy			
0:with	1.210	.010	.001
1:without			
Mental Activities			
0:with	.579	.000	.003
1:without			
Physical Activities			
0:with	.546	.002	.000
1:without			
Blood Pressure			
0:with	1.079	.651	.780
1:without			

The log rank test is testing the null hypothesis that there is no difference in the overall survival distributions between the groups (e.g., intervention groups) in the population. To understand how the survival distributions compare between groups, Figure of the cumulative survival proportion against time for each intervention group are presented as the Survival Functions plot.

The regression coefficients provide us to quantify the log of the hazard in the treatment groups, based on the covariates included in the model; it is interpreted as a relative risk (assuming no time-varying coefficients). To investigate the effects of individual predictors, Exp(B) is interpreted, which is the hazard ratio and can be interpreted as the predicted change in the hazard for a unit increase in the predictor. A value of < 1 says that an increase in one unit for that particular variable will decrease the probability of experiencing an end point throughout the observation period. By inverting (that is $1/Exp(B)$), one may find the “protective effect”.

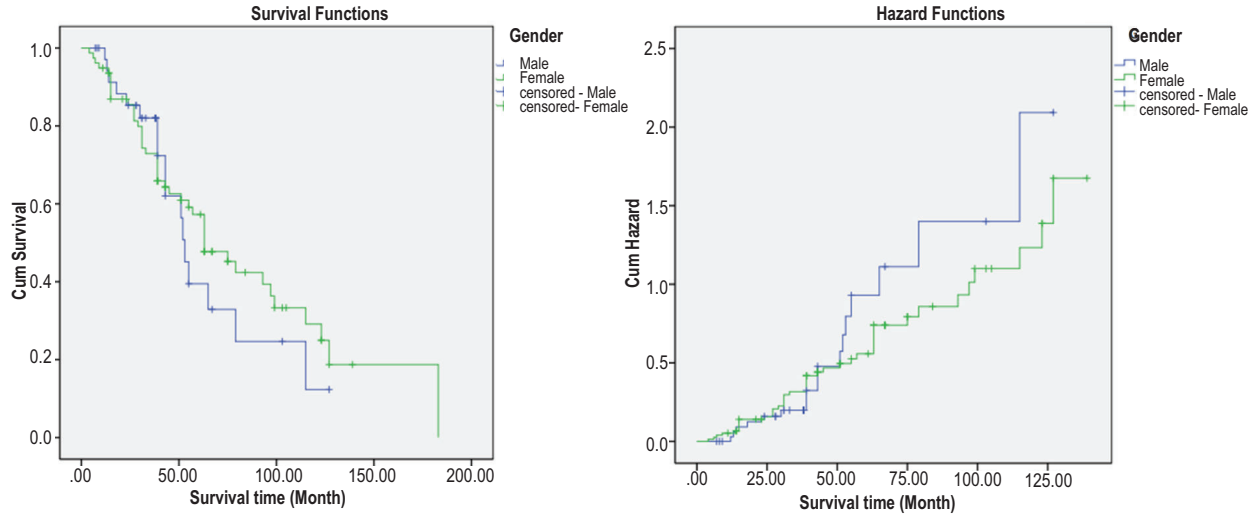


Figure 1: *Survival and Hazard Functions for Gender*

As the Figure 1 reveal the Survival rate for Females is higher than men after 50 months, while consequently the Hazard Function for Men is dramatically higher, which do appear to differ considerably.

Table I displays that the Log Rank test *p*-value for Gender is actually .00 (i.e., *p* = .02), have a statistically significant result and ONE can conclude that the survival distributions of the

Age has been categories in 3 different stages, before 75, between 75 and 89 and older than 89. Survival functions in Figure 2 present decreasing pattern except for patients younger than 75 which is constant after 40 months. Plus the Hazard function appears to be increasing with a constant pattern in the same situation. Based on Table I, Log Rank test *p*-value for Age is **.001**; one can reach to the fact that the survival distributions of the different categories of Age are not equal in the population. The HR for Age is

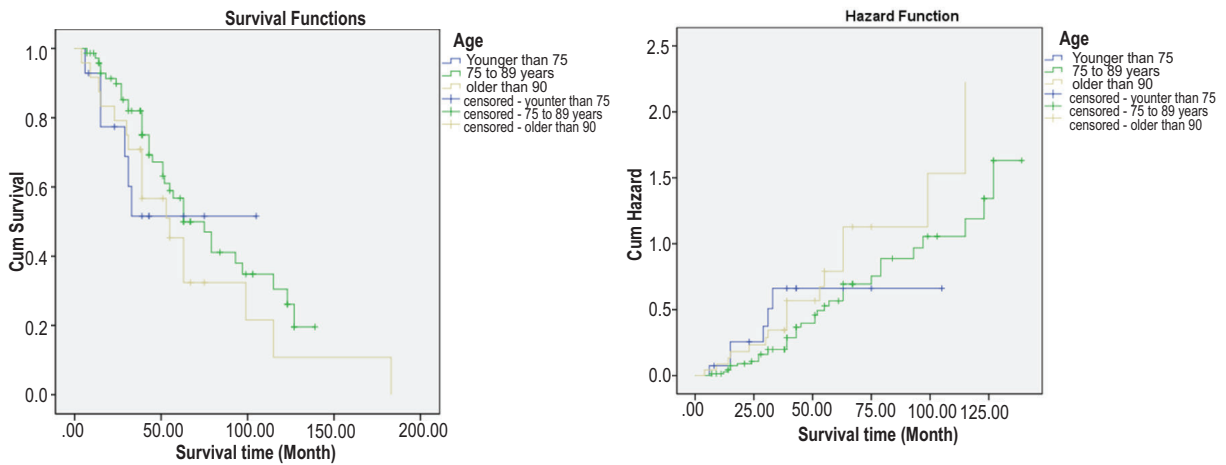


Figure 2: *Survival and Hazard Functions for Age*

different Genders of intervention are not equal in the population. Based on $\text{Exp}(B)=1.72$ (Sig=.000) , the interpretation would be that Hazard ratio for men is greater than women, basically having the value of Gender = 1 (Female) means that you increase the probability of experiencing an end point with $1/1.72=.58$, compared to when the Gender value = 0(Male).

1.19(Sig=.001), so an increase in Age of 1 year will be associated with a 1.19 unite increase in the hazard.

We can see from Table I that the P-Value and also Log Rank test for Addiction are not significant, leads to same survival and Hazard rate in addicted and non- addicted patients, thus the Factor Addiction presents no significant effect in our population-based study.

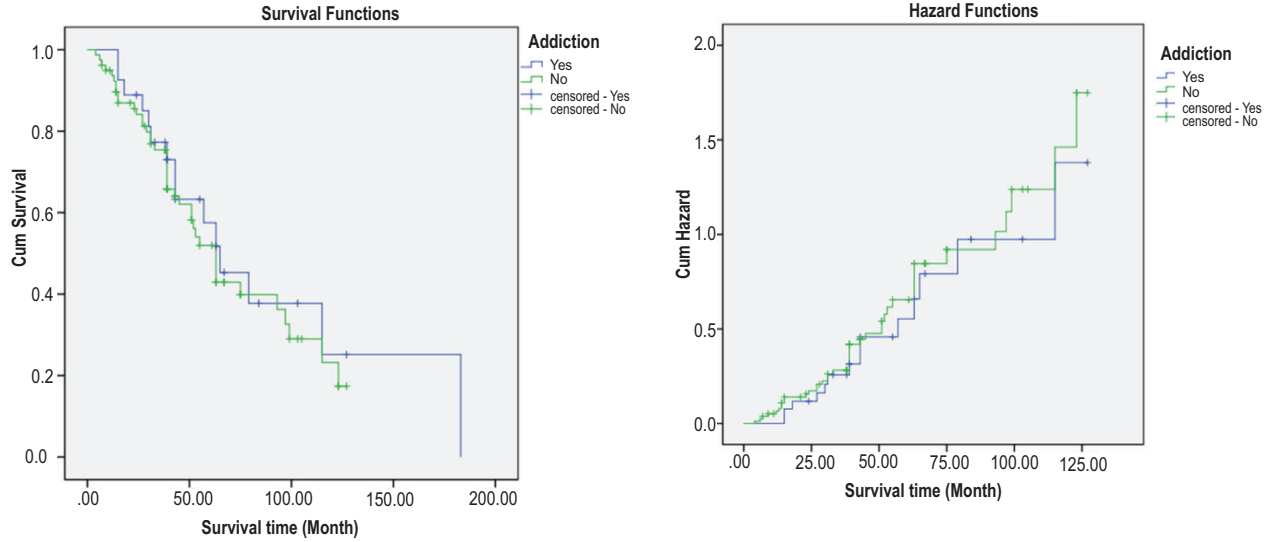


Figure 3: Survival and Hazard Functions for Addiction

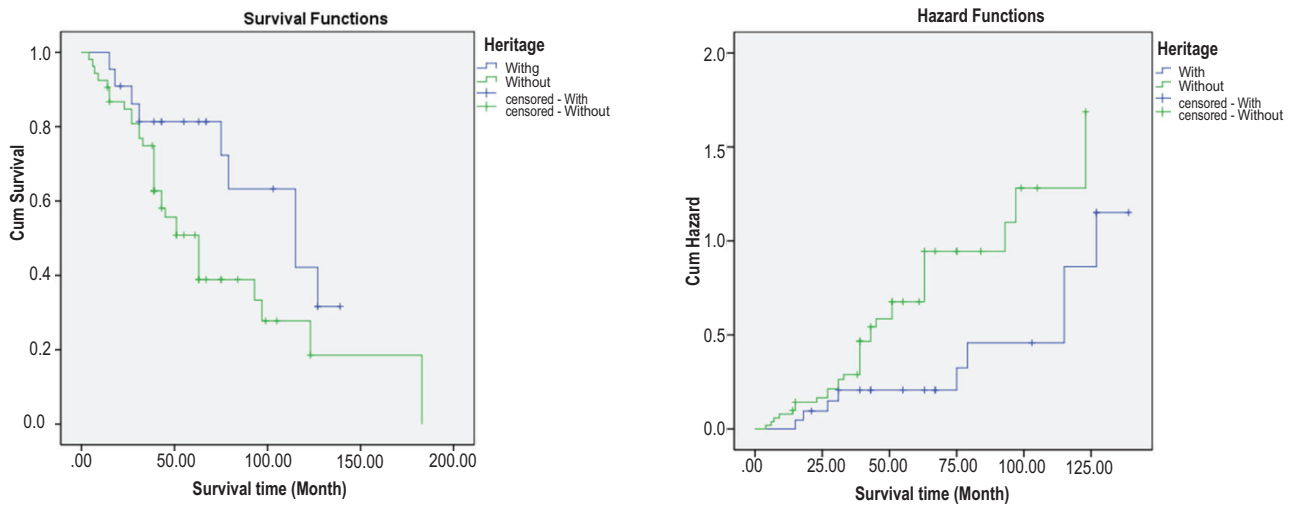


Figure 4: Survival and Hazard Functions for Heritage

As the Figure 4 reveal the Survival rate for patients with AD inherited in the family is higher than patients without hereditary. Consequently the Hazard Function for the first group is dramatically higher.

Table I displays that the Log Rank test p -value for Heredity is actually .00 have a statistically significant result and can conclude that the survival distributions of the different group of intervention do differ significantly in the population. Based on $\text{Exp}(B)=.451$ ($\text{Sig}=.012$), the interpretation will be that having the value of Heredity = 1 means that one may decrease the probability of experiencing

an end point with $1/0.451=2.21$, compared to when the Heredity value = 0.

The next Figure show that the survival probability is lower for patient with Apoplexy at all-time points so they are less likely to survive. Based on Table I, Log Rank test p -value for Apoplexy is .010; one can conclude that the survival distributions of the different categories of Apoplexy are not equal in the population. Based on $\text{Exp}(B)=1.21$ ($\text{Sig}=.001$), means that scoring “Apoplexy” = 1 results in an increased probability of experiencing an end point compared to when “Apoplexy” = 0.

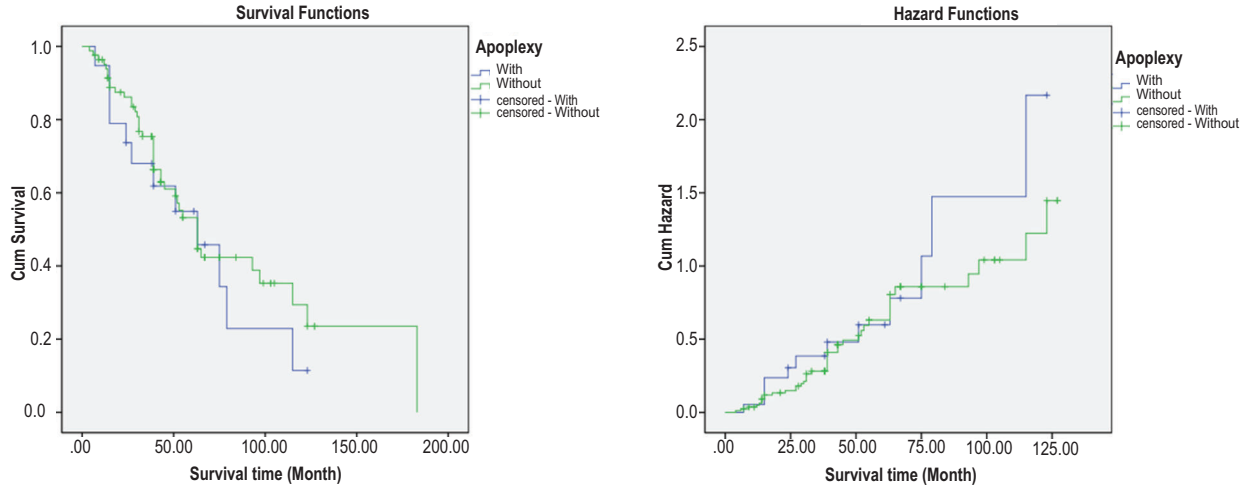


Figure 5: Survival and Hazard Functions for Apoplexy

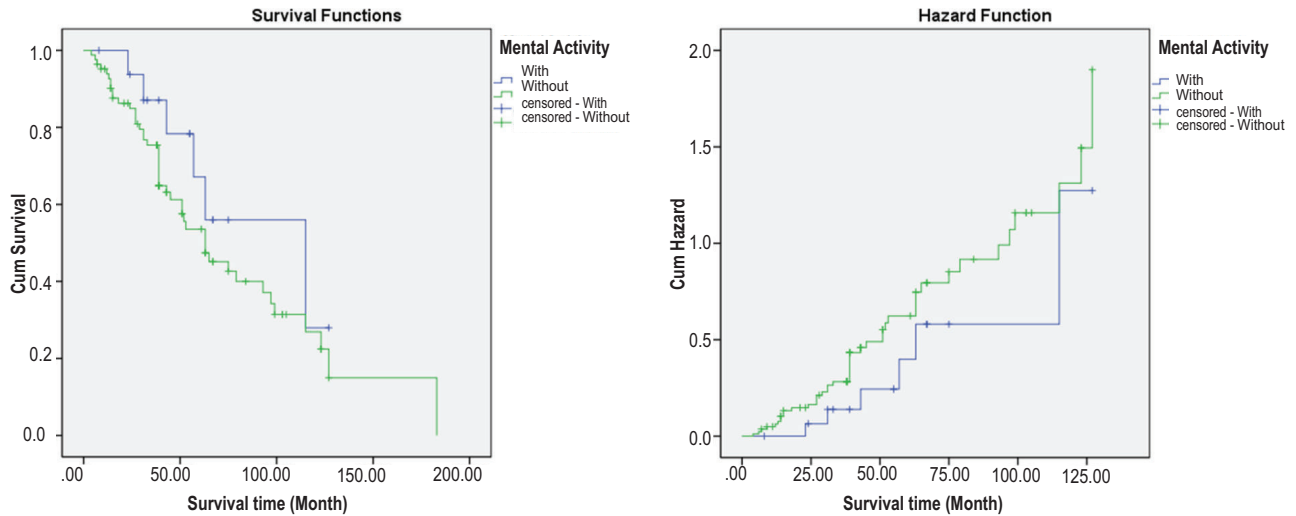


Figure 6: Survival and Hazard Functions for Mental Activity

As the Figure 6 reveal the Survival rate for patient with Mental Activity is higher than patient without Mental Activity, while consequently the Hazard Function is dramatically higher for the second group.

Table I provides us the Log Rank test p -value for Mental Activity is actually $p=.000$, have a statistically significant result and can conclude that the survival distributions of 2 grouping of intervention are not equal in the population. Based on $\text{Exp}(B)=.579$ ($\text{Sig}=.003$), the interpretation will be that having the value of Mental Activity = 1 means that the probability of experiencing an end point with $1/.579=1.72$

would be dramatically decreased, compared to when the Gender value = 0.

The last Figure shows that the survival probability is lower for patient without Physical Activities at all-time points so they are less likely to survive. Based on Table I, Log Rank test p -value for Apoplexy is .002; one can conclude that the survival distributions of the different categories of Physical Activities are not equal in the population. Based on $\text{Exp}(B)=.546$ ($\text{Sig}=.000$), the group with Physical Activates will result in an decreased probability of experiencing an end point compared to the group without Physical Activities.

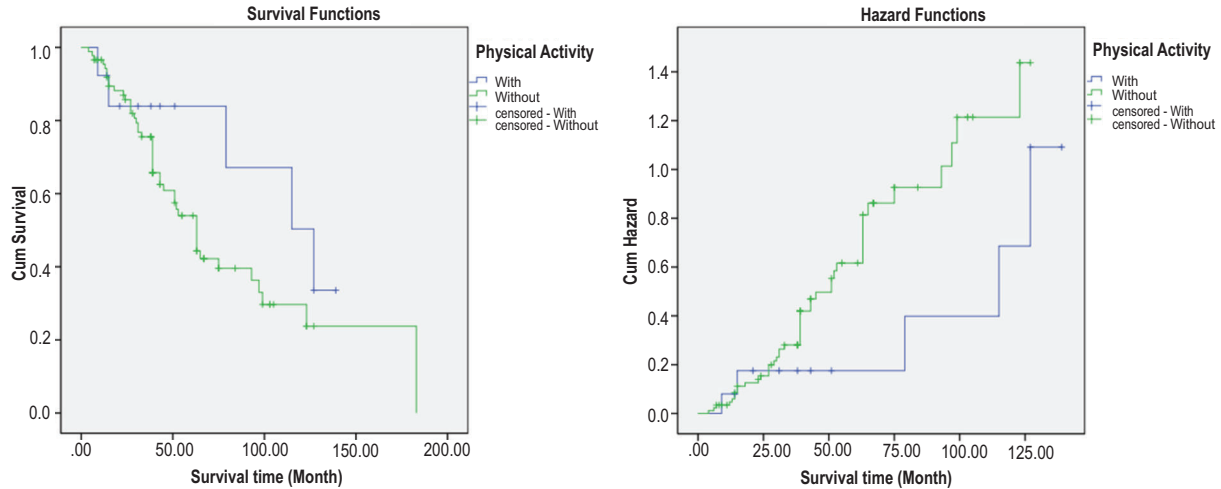


Figure 7: Survival and Hazard Functions for Physical Activity

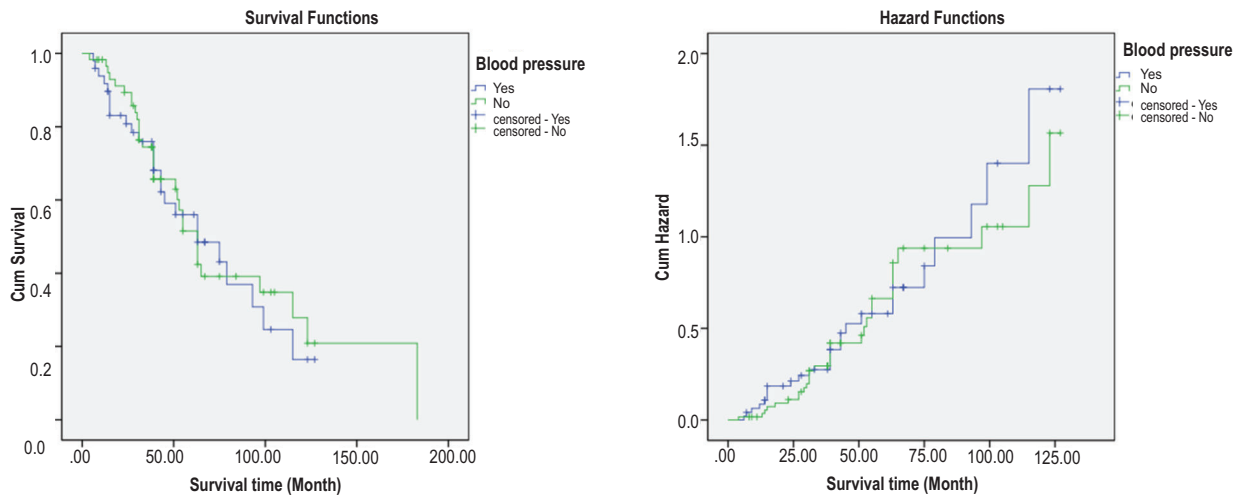


Figure 8: Survival and Hazard Functions for Blood pressure

We can see from Table I that the P-Value and also Log Rank test for Blood pressure are not significant, leads to same survival and Hazard rate in 2 mentioned groups.

Discussion:

This study is the first attempt to estimate the prevalence of AD in Iran using individual data. The number of patients with Alzheimer disease is expected to grow in modern industrial status in Iran, like Tehran and also Khorasan States. This paper discusses the modifiable risk factors associated with Alzheimer disease risk in mentioned 2 states based on Cox Proportional Hazard models. Cox’s regression model is considered to be semi-parametric. The model is able to consider all different hazard functions when predicting the hazards, but does not rely on any particular distribution.

Factors influencing survival in Alzheimer disease include Age (Sig=.001), Gender (Sig=.020) and Family story (Sig=.000). Moreover, Life style habits which include Mental (Sig=.000) and physical (Sig=.002) activities appear to have positive effect on modifying Survival function, i.e., our population-based studies reported the positive association between Life style habits and Survival time. In contrast, we found no link between survival duration and a history of Addiction and Blood Pressure in our population-based study.

Conclusions:

Based on mentioned results, since Tehran and Khorasan have experienced modern and industrial life styles during 30 years, government needs to plan and enforce programs for the high-risk groups and also conduct more vast researches in the regional and national levels.

Conflict of interest: None.

References:

1. Duron E, Hanon O. Hypertension, cognitive decline and dementia. *Arch Cardiovasc Dis.* 2008;101(3):181–9.
2. Geldmacher DS. Differential diagnosis of dementia syndromes. *Clin Geriatr Med.* 2004;20(1):27–43. doi: 10.1016/j.cger.2003.10.006.
3. Flicker L. Modifiable lifestyle risk factors for Alzheimer's disease. *J Alzheimers Dis.* 2010;20:803–811.
4. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol.* 2011;10(9):819–28.
5. Chen JH, Lin KP, Chen YC. Risk factors for dementia. *J Formos Med Assoc.* 2009;108(10):754–64.
6. Razay G, Vreugdenhil A, Wilcock G. Obesity, abdominal obesity and Alzheimer disease. *Dement Geriatr Cogn Disord.* 2006;22(2):173–6.
7. Katzman R. Editorial: The prevalence and malignancy of Alzheimer disease. A major killer. *Arch Neurol.* 1976;33(4):217–8.
8. Kennelly SP, Lawlor BA, Kenny RA. Blood pressure and the risk for dementia: a double edged sword. *Ageing Res Rev.* 2009;8(2):61–70.
9. Nozari N, Ferri CP, Farin F, Noroozian M, Salehi M, Seyedian M, et al. Validation of the 10/66 Dementia Research Group's 10/66 dementia diagnosis in Iran. *Int Psychogeriatr.* 2009;21:604–605.
10. Hall CB, Verghese J, Sliwinski M, Chen Z, Katz M, Derby C, et al. Dementia incidence may increase more slowly after age 90: results from the Bronx Aging Study. *Neurology.* 2005; 65(6):882–6. doi: 10.1212/01.wnl.0000176053.98907.3f.
11. Singer, Judith D.; Willett, John B. (2003). "Fitting Cox Regression Models". *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* New York: Oxford University Press. pp. 503–542.
12. Therneau, T. M.; Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model.* New York: Springer.
13. Sosa-Ortiz AL, Acosta-Castillo I, Prince MJ. Epidemiology of dementias and Alzheimer's disease. *Arch Med Res.* 2012;43(8):600–8.