Considering time-interaction terms using parametric survival models for interval-censoring data

Erfan Ghasemi [1], Alireza Akbarzadeh Baghban [2], Ahmadreza Baghestani [3], Saeed Asgary [4]

(1) Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
(2) Physiotherapy Research Centre, School of Rehabilitation, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
(3) Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
(4) Iranian Center for Endodontic Research, Institute of Dental Research, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

CORRESPONDING AUTHOR: *Alireza Akbarzadeh Baghban, Tel: 0098-912-2955411; Fax: 0098-21-22707347;E-mail: akbarzad@gmail.com; Address: Damavand St. School of Rehabilitation, Shahid Beheshti University of Medical Sciences, Tehran, Iran.*

DOI: ***10.2427/12134***
Accepted on February 27, 2017

ABSTRACT

Introduction: Many of the variables considered for studying in survival research are time-invariant, i.e. their values do not change over time whiletheir effects may vary over time. Therefore, the change in behavior taking place over time needs to be included in the analysis, which is done by adding time-interaction terms to the model.
Method: In this research, a parametric survival model, which is able to evaluate the effect of time-dependent variables, was applied for interval-censored data in such a way that the time for invariant variables interaction was considered as time-dependent variables.
Result: Using a practical example, the results indicated that parametric survival model can alter the interpretations regarding the effects of exploratory variables.
Conclusion: Regarding fixed variables whose effects change over time, the researcher can incorporate their interaction effect with time, and treat them as time-dependent variables and obtain appropriate inferences.

Key words: Survival Analysis, Parametric Models, Interval-Censoring, TimeDependent Variables

INTRODUCTION

In medical and biological research, survival analysis has been developed for investigating the effect of covariates on the risk function of a population.

Censoring is an important feature which differentiates survival analysis from other analyses in statistics [1]. According to [2], censoring is related to the situation in which the event times are not observed well. Based on different situations, there are different kinds of censoring such as right-censored data, left-censored data, and interval-censored. Right censoring takes place when a subject leaves the study before an event , or the study ends before the occurrence of an event. Left censoring means that the event of interest has already taken place before enrolment. Regarding interval censoring, the event of interest cannot be directly observed and it is only recognized during a random interval of time [1]. In interval censoring, both the beginning and the ending time are considered. Therefore, the exact time of the event happens between these two times. Censoring may take place in various ways, especially when the survival

times are discrete. Therefore, a variety of parametric and semi-parametric models were developed for such interval-censored data. In some research, semi-parametric modeling was suggested for interval censored data [3-6]. Also, some focused on parametric models [7, 8]. Younes and Lachin presented a link-based model that could be applied to the interval-censored case [9]. Finally, Sparling et al. [10] proposed a group of parametric survival models for left, right, and interval-censored data with fixed and time-dependent covariates.

In survival analysis, many of the considered variables are time-invariant. In fact, their values do not change over time althoughtheir effects may change over time in some cases. The change of effect during the time should be considered in analysis and this can be done only when time-interaction variables included in the model.

In the present study , time-interactions are defined as the variables created by the product of the original variables during the time so that they can be considered as time-dependent variables. In other words, according to Sparling et al. [10], the parametric survival models were applied, which allow the researcher to include time-interaction instead of time dependent covariate in the models. Using these models bring encompass some advantages. First, these models have some special features such as proportional hazard and proportional odds models. Furthermore, when the data are heavily interval-censored, the parametric models yield more robust and informative results than their corresponding semi-parametric models [11].

METHODS

In general, there are three types of censoring survival data including right-censored, left-censored and interval-censored. To distinguish among the three types of censors, the indicator functions are defined. For the i^{th} subject (i = 1,2, …, n), x_i is the event time, and t_i is the variable for alternatively representing the observed time event, right-censored or left-censored time. In interval-censored, t_{Li} is the left-censored time, t_{Ri} is the right-censored time and $t_{Li} < x_i < t_{Ri}$ is the interval-censored time.

$$\delta_{R_i} = \begin{cases} 1 & \text{if right censored at time } t_i < x_i \\ 0 & \text{Other wise} \end{cases}$$

$$\delta_{L_i} = \begin{cases} 1 & \text{if left censored at the time } t_i > x_i \\ 0 & \text{Other wise} \end{cases}$$

$$\delta_{I_i} = \begin{cases} 1 & \text{if interval censored with } t_{L_i} < x_i < t_{R_i} \\ 0 & \text{Otherwise} \end{cases}$$

$$\delta_{E_i} = \begin{cases} 1 & \text{if has an event observed exactly at } t_i = x_i \\ 0 & \text{Otherwise} \end{cases}$$

Then, $\delta_{E_i} = 1 - (\delta_{R_i} + \delta_{L_i} + \delta_{I_i})$.

Odell et al [7] defined a likelihood function by considering right-censored, left-censored and interval censored survival data. To define the likelihood function, suppose that f(t) represents the probability for density function of event time ,and F(t) denotes the cumulative distribution function. Based on the assumption that the censors and the events are independent of each other, theb likelihood function for each sample of n independent subjects is as follows:

$$L = \prod_{i=1}^{n} [f_i(t_i)^{\delta_{E_i}} F_i(t_{L_i})^{\delta_{L_i}} [1 - F_i(t_{R_i})]^{\delta_{R_i}} [F_i(t_{R_i}) - F_i(t_{L_i})]^{\delta_{I_i}}] \quad (1)$$

When (p) is considered as the number of constant variables and (q) as the number of time-dependent variables in a model, the equation $z_i = (z_{i1}, \ldots, z_{ip})$ is a fixed covariate vector for the ith subject. Now, suppose that for the ith subject, time-dependent covariates have changed over time $(\tau_{i0}, \ldots, \tau_{ikj})$, and new values are replaced by them (τ_{i0} is the time that the i^{th} subject enters the study) while time sets of $\{\tau_{ij}\}$ might be different for different subjects. Also, suppose that $y_i = (y_{ij1}, \ldots, y_{ijq})$ is a vector of q time dependent variables for the ith subject at the time of τ_{ij} which is shown by $y_i(\tau_{ij})$, and $y_i[\tau_{ikj}] = (y_{i0}, \ldots, y_{ikj})$ is a complete sequence of time-dependent covariates of the study for the i^{th} subject. Based on this definition, $y_i[\tau]$ represents a sequence of time-dependent covariate values up to time t for the i^{th} subject. Under this model, covariates change at discrete times. Therefore, the likelihood function is as follows [10]:

$$L = \prod_{i=1}^{n} (f_i(t_i|z_i, Y_{i[(t_{ij})]})^{\delta_{E_i}} F_i(t_i|z_i, Y_{i[(t_{ij})]})^{\delta_{L_i}} \times$$

$$[1 - F_i(t_i|z_i, Y_{i[(\tau_{ij})]})]^{\delta_{R_i}} [F_i\left(t_{R_i}|z_i, Y_{i[(\tau_{R_i})]}\right) -$$

$$F_i(t_{L_i}|z_i, Y_{i[(t_{L_i})]})]^{\delta_{I_i}}\}$$

(2)

Where, $f_i(t_i|z_i, y_{i[(\tau i)]})$ is the event density for an event time conditional distribution condition on values of fixed covariates z_i, and the sequence of time-dependent covariate values for a subject.

Suppose γ and η are the vectors of coefficients for fixed and time-dependent covariates z_i and y_{ij}:

$$\gamma' z_i = \gamma_1 z_{i1} + \cdots + \gamma_p z_{ip}$$

$$\eta' z_i = \eta_1 y_{i1} + \cdots + \eta_q y_{iq}$$

In the present study, Weibull distribution was used. The density function of Weibull distribution is as follows [12]:

$$f(t) = \lambda p t^{p-1} \exp(-\lambda t^{p})$$

The density function was calculated at each time and was rewritten for each subject.Therefore, we have the following equations:

$$\lambda_{ij} = \exp(\beta_0 + \gamma' z_i + \eta' z_i) \quad (3)$$

where, β_0 is intercept. Regarding the specific distribution of the data and its status in terms of the type of censorship or observed event, the defined likelihood function can be established by estimating the parameters and identifying the effective variables on survival time of the patients under study.

Applied example

The applied example is a five-year multi-center randomized clinical trial (RCT) in the field of dentistry. The dataset consisted of 246 women and 144 men and the trial was approved and assessed by the Iranian Ministry of Health as well as the Ethics Committee of the Dental Research Center of Shahid Beheshti University of Medical Sciences, Tehran, Iran. The purpose of the RCT was to compare the existence of periapical lesion after one-visit therapy (OET) and a performer with a new endodontic biomaterial (CEM cement; PCC) in human permanent molars with irreversible pulpits. A total of 390 selected patients who met the inclusion criteria were randomly entered into the OET group (n =195) and the PCC group (n =195). These patients were selected from 23 health care centers in four states and five medical universities of Iran. All the treated teeth were clinically followed up after pulpotomy or endodontic therapy (baseline 6th, 12th, 24th and 60th month).The response variables in this example were either the periapical lesion (failure), or no periapical lesion (Success). The predictors in this model included an indicator of gender (0=female, 1=male), age (years), and that of treatment type (PCC or OET) [13].

RESULTS

The patients had a mean age of 26.26 in OET group while it was 26.63in the PCC group. Therefore, there was no significant difference between the two groups in based on age (P=0.651). The number of females in in OET group was 118 subjects (60.5%) and 128 (65.6%) in PCC group (P=0.294).

Simple OR calculations were used for data analysis. Table 1 shows the odds ratio (OR) of success in PCC as the treatment group, compared with the OET, at different follow-up times.

The results indicated that there is a decrease in the odd ratio over time asthe success odds at 6th and 12th month were significantly higher in PCC group, compared to those in the OET group. However, no significant difference was reported between the two group at 24th and 60th month.

Figure 1 illustrates the survival curves by using Turnbull empirical estimates, whichis used in interval censoring [14]. As it is evident from this figure, the two curves get closer to each other after the 12th month.

Table 2 indicates Maximum likelihood estimation of regression coefficients and Wald test results for ordinary and time-dependent Weibull models. Given Weibull distribution, interpretation of proportional hazard regression coefficients is based on log hazard ratio or log relative risk.

Results indicated that the estimated coefficients in both models had similar values for both gender and age variable and there was no significant difference between these two variables in the models.

The interaction coefficient between group and time in time-dependent Weibull model was significant as the interaction coefficient and the treatment group had two opposite signs. In other words, the hazard in OET and PCC get closer to each other over time or the treatment in PCC group gradually loses its superiority over the OET group after some time. Therefore, we can concluded that the ordinary Weibull model fails to estimate the interaction effect between group and time.

The coefficient of treatment group was estimated to be 0.429,based on ordinary Weibull model. That is, hazard of failure in OET group is 1.5 times (exp (0.429) = 1.53) higher than the PCC group. Based on the time-dependent Weibull model, the proportional hazard is obtained separately at each time interval. For example, after one-year treatment, hazard in OET group becomes 2.72 times greater than PCC group (exp(1.261-0.261*time)=2.72) while there is a 2.01

TABLE 1. Odds ratio of success to failure in the treatment group PCC compared with the OET

Time (month)	OR	95% Confidence Interval for OR	
		Lower	Upper
6	3.603	1.980	6.556
12	3.206	1.632	6.300
24	1.507	0.839	2.707
60	1.324	0.811	2.161

FIGURE 1. Survival curves and confidence band for the two treatment groups using Turnbull empirical estimates.

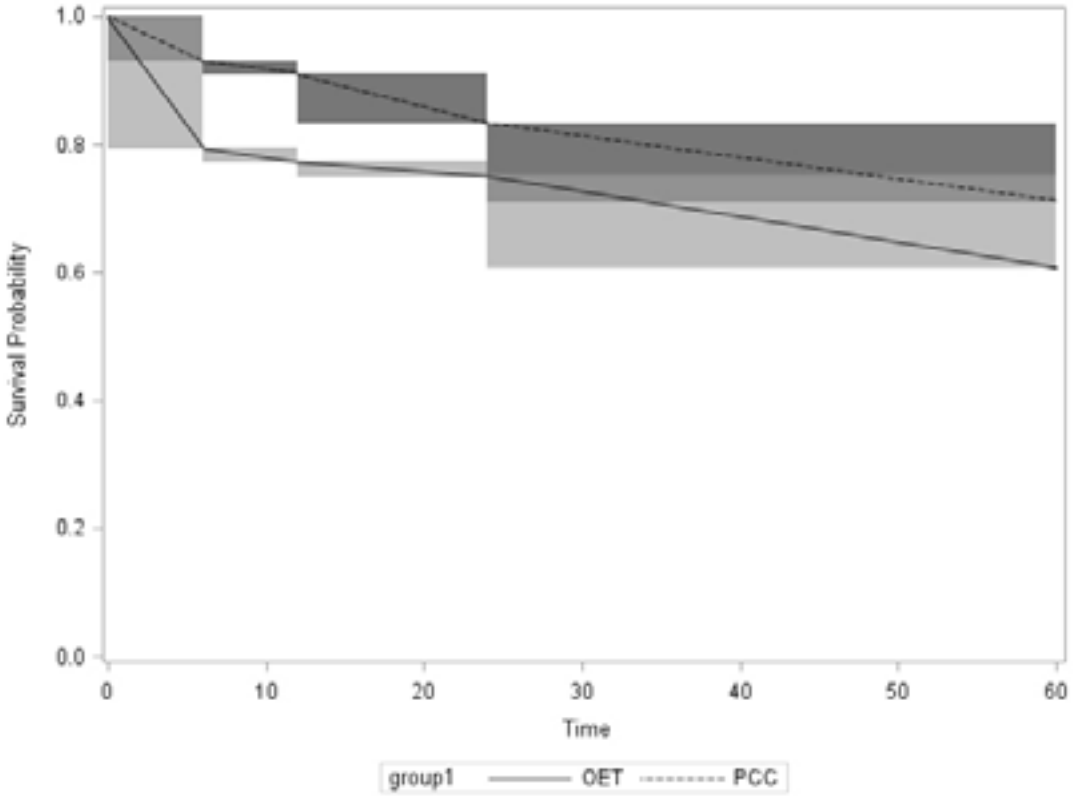

TABLE 2. Maximum likelihood estimation of regression coefficients and Wald test for ordinary and time-dependent Weibull model.

	ORDINARY WEIBULL MODEL			TIME-DEPENDENT WEIBULL MODEL		
	Estimate	**Std. Err.**	**P-value**	**Estimate**	**Std. Err.**	**P-value**
Intercept	-2.348	0.388	<0.001*	-3.379	0.477	<0.001*
Gender[a]	0.201	0.187	0.283	0.215	0.187	0.251
Age	-0.015	0.012	0.211	-0.015	0.012	0.183
Treatment Group[b]	0.429	0.186	0.021*	1.261	0.370	0.001*
Time*Group[c]	–	–	–	-0.261	0.099	0.008*
Shape parameter	0.475	0.056	–	0.637	0.091	–

**: Significant at 0.05 level; [a]: Female is reference group; [b]: PCC is reference group; [c]: Scale of time variable is years*

decrease in the ratio after two-year treatment (exp(1.261-0.261*2)=2.01).

CONCLUSION

The present study aimed to examine the interaction effect of fixed and time variables, based on time-dependent Weibull model , which was suggested by Sparling et al. to investigate the effect of time-dependent variables on survival time. In order to make a comparison, ordinary Weibull model was also implemented.

The fitting results were similar for the fixed variables in both models while different results were obtained for the variable whose effect changes over time although it was significant in both models. In time-dependent Weibull model, time interaction is also incorporated as a predictor, which is able to demonstrate behavior change in the variable over time.

Based on the results , the odds ratio of success in PCC group decreases over time, compared to the OET group (Table 1). The results of time-dependent Weibull survival model confirmed that the hazard ratio decreased in OET group, compared to PCC group, over time. Akbarzade et al. investigated the data related to the two-year follow-up of the clinical study by using longitudinal models [15].

The results indicated that PCC group had higher chance to success, compared to OET group. The results are consistent with the findings of the present study althoughthe therapy effect of both treatment methods becomes similar over time in the present study.

In order to analyze interval censored survival times, various parametric, semi-parametric and nonparametric models were suggested [4, 11, 16-18]. Further, some models have been proposed for analyzing time-dependent variables [19, 20]. The model used in the present study considered interval-censoring and adjusted the time-dependent variable effects. This model is applied to those variables that are seemingly fixed but their effects change over time. Further, the interactions among the variables with time are taken into account. According to Sparling, the inferences obtained from this model are based on large sample approximations; therefore, certain precautions should be taken in dealing with the results related to small sample sizes.

In conclusion, regarding fixed variables whose effects change over time, the researcher can incorporate their interaction effect with time, and treat them as time-dependent variables and obtain appropriate inferences. Here, it was assumed that the response variable is Weibull distribution, which is an appropriate distribution in many survival studies. As for other distributions, other models such as log logistic can be applied for time-dependent variables with interval-censored data.

References

1. Gómez G, Calle ML, Oller R. Frequentist and Bayesian approaches for interval-censored data. Stat. Pap. 2004; 45(2): 139-73.
2. Remontet L, Bossard N, Belot A, Esteve J. An overall strategy based on regression models to estimate relative survival and model the effects of prognostic factors in cancer survival studies. Stat. Med. 2007; 26(10): 2214-28.
3. Finkelstein DM, Wolfe RA. A semiparametric model for regression analysis of interval-censored failure time data. Biometrics. 1985;41: 933-45.
4. Finkelstein DM. A proportional hazards model for interval-censored failure time data. Biometrics. 1986; 42: 845-4.
5. Seaman S, Bird S. Proportional hazards model for interval censored failure times and time dependent covariates: application to hazard of HIV infection of injecting drug users in prison. Stat. Med. 2001; 20(12): 1855-70.
6. Betensky RA, Lindsey JC, Ryan LM, Wand M. A local likelihood proportional hazards model for interval censored data. Stat. Med. 2002; 21(2): 263-75.
7. Odell PM, Anderson KM, D'Agostino RB. Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. Biometrics.1992; 48: 951-9.
8. Rabinowitz D, Tsiatis A, Aragon J. Regression with interval-censored data. Biometrika. 1995; 82(3): 501-13.
9. Younes N, Lachin J. Link-based models for survival data with interval and continuous time censoring. Biometrics. 1997; 53: 1199-211.
10. Sparling YH, Younes N, Lachin JM, Bautista OM. Parametric survival models for interval-censored data with time-dependent covariates. Biostatistics. 2006; 7(4): 599-614.
11. Lindsey J. A study of interval censoring in parametric regression models. Lifetime Data Anal. 1998; 4(4): 329-54.
12. Kleinbaum DG, Klein M. Survival analysis. Springer, 1996.
13. Asgary S, Eghbal MJ, Fazlyab M, Baghban AA, Ghoddusi J. Five-year results of vital pulp therapy in permanent molars with irreversible pulpitis: a non-inferiority multicenter randomized clinical trial. Clin. Oral Investig. 2014; 19(2): 335-41.
14. Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. J. Roy. Stat. Soc. B Met. 1976; 290-5.
15. Baghban AA, Ghasemi E, Zayeri F, Asgary S, Namdari M. Analysis of longitudinal binary outcomes in clinical trials with low percentage of missing values. J. Paramed. Sci. 2013; 4(Supp): 33-41.
16. Sinha D, Dey DK. Semiparametric Bayesian analysis of survival data. J. Am. Stat. Assoc. 1997; 92(439): 1195-212.
17. Goetghebeur E, Ryan L. Semiparametric Regression Analysis of Interval Censored Data. Biometrics. 2000; 56(4): 1139-44.
18. De Gruttola V, Lagakos SW. Analysis of doubly-censored survival data, with application to AIDS. Biometrics. 1989; 45(1): 1-11.
19. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. Annu. Rev. Publ. Health. 1999; 20(1): 145-57.
20. Zucker DM, Karr AF. Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. Ann. Stat. 1990; 18: 329-53.

*