RESEARCH ARTICLE

F.U. Med.J.Health.Sci.
2024; 38 (3): 261 - 268
http://www.fusabil.org

# Model Performance Evaluation in Clustered Survival Data: A Simulation Study*

**Kübra Elif AKBAŞ**[1, a]
**Harika Gözde GÖZÜKARA BAĞ** [2, b]

[1] Firat University,
Faculty of Medicine,
Department of Biostatistics and Medical Informatics,
Elazig, TURKIYE

[2] İnönü University,
Faculty of Medicine,
Department of Biostatistics and Medical Informatics,
Malatya, TURKIYE

[a] ORCİD: 0000-0002-2804-000X
[b] ORCİD: 0000-0003-1208-4072

**Objective:** Clustered data structure is a frequently encountered data type today. Like the other analysis when performing a survival analysis, clustered data type should be taken into account. The aim of this study is to compare some survival analysis used for clustered data and Cox regression analysis.

**Material and Method:** The study consists of three parts. In the scenarios in the first section, cluster size and number of individuals change in balanced cluster sizes. In the second part, the effect of changes in the number of individuals in unbalanced cluster sizes, and in the third section, only the effect of changing censor rates while the cluster sizes and number of individuals remained constant were examined. In this study, 5 different models were used to apply simulated data and AIC, AICc and BIC were used to compare their performances.

**Results:** Within the scope of the findings, the best model based on AIC and AICc is the frailty model. According to the BIC, the best model was obtained as the marginal Cox model. In the simulation studies, the worst model changes in parallel with the increase in the sample size.

**Conclusion:** As a result, ignoring the data structure can lead to biased or inaccurate estimates, especially in health data. In this study, it is recommended to use the frailty or marginal Cox model in clustered survival data depending on the information criteria.

***Key Words:*** *Survival anaylsis, clustered survival data, Cox model, condinitional models, marginal models*

## Kümelenmiş Sağkalım Verilerinde Model Performans Değerlendirmesi: Bir Simülasyon Çalışması

**Amaç:** Kümelenmiş veri yapısı günümüzde sıklıkla karşılaşılan bir veri türüdür. Hayatta kalma analizi yapılırken diğer analizlerde olduğu gibi kümelenmiş veri türü dikkate alınmalıdır. Bu çalışmanın amacı kümelenmiş veriler için kullanılan bazı sağkalım analizleri ile Cox regresyon analizini karşılaştırmaktır.

**Gereç ve Yöntem:** Çalışma üç bölümden oluşmaktadır. Birinci bölümdeki senaryolarda dengeli küme büyüklüklerinde küme boyutu ve birey sayısı değişmektedir. İkinci bölümde dengesiz küme büyüklüklerinde birey sayılarının değişimlerinin etkisi, üçüncü bölümde ise küme büyüklükleri ve birey sayıları sabit kalırken sadece sansür oranları değişiminin etkisi incelenmiştir. Bu çalışmada simüle edilmiş verilerin uygulanması için 5 farklı model kullanılmış ve performanslarının karşılaştırılmasında AIC, AICc ve BIC kullanılmıştır.

**Bulgular:** Bulgular kapsamında AIC ve AICc' e göre en iyi model kırılganlık modelidir. BIC' e göre ise en iyi model marjinal Cox modeli olarak elde edilmiştir. Simülasyon çalışmalarında örneklem büyüklüğünün artmasına paralel olarak en kötü model değişmektedir.

**Sonuç:** Sonuç olarak veri yapısının göz ardı edilmesi, özellikle sağlık verilerinde taraflı veya hatalı tahminlere yol açabilir. Bu çalışmada bilgi kritetinede bağlı olarak kümelenmiş sağkalım verilerinde kırılganlık veya marjinal Cox modelinin kullanılması önerilmektedir.

**Anahtar Kelimeler***: Sağkalım analizi, kümelenmiş sağkalım verisi, Cox model, koşullu modeller, marjinal modeller*

**Received** : 21.05.2024
**Accepted** : 22.09.2024

**Correspondence Address**

**Kübra Elif AKBAŞ**
Firat University,
Faculty of Medicine,
Department of Biostatistics and Medical Informatics,
Elazig, TURKIYE

**keet@firat.edu.tr**

## Introduction

Survival analysis is used in a variety of fields to analyze data that includes the time between two events or, more generally, the transition times between several states or conditions (1). This analysis is concerned with examining survival times and the factors affecting them. Survival analysis study types include clinical trials, prospective and retrospective observational studies, and animal experiments (2). The most important feature that distinguishes survival analysis from other statistical analysis is that survival data is often censored or incomplete in some way (1, 3, 4).

In some applications, data may be clustered or correlated due to some common characteristics such as genetic characteristics or shared environmental factors. Clustered survival data are used to represent observed multivariate data with naturally formed clusters and parallel event times for each cluster (5). When an identical event

* This paper was produced from PhD thesis submitted to İnönü University by Kübra Elif AKBAŞ.

type is evaluated on two or more separate, comparable units within a same subject, the resulting clustered survival data is produced. Or, multivariate survival data emerge when different events or repeated events occur in the same individual (6).

In many applications in survival analysis, data are organized at clusters with which survival times can be correlated (7). Observations from the same cluster often share some unobserved features and tend to be correlated as a result. Although most researchers agree that this should be considered in the analysis of survival times, it is often ignored. This may lead to inadequate and biased estimates. Because of the predictably more serious consequences, the precision of the estimates is likely to be overstated if observations in the same cluster are positively correlated (8). Clustered data appear frequently in the medical field. One of these areas is periodontal studies. Multiple areas (gums or teeth) in each individual's mouth are examples of clustered structure (9).

Tooth decay is a common disease that is and continues to be the most common chronic disease in the modern world. It is defined as a chronic dynamic process or disease that occurs in the tooth structure when in contact with bacterial plaque accumulation, especially microbial deposits (10). Tooth decay and chronic gingivitis can occur due to bacterial plaque on and around the teeth (11). Since each individual has more than one tooth and has different susceptibility to dental caries, it is very important to evaluate the formation of dental caries at the individual person level as the unit of analysis in survival methods (12). In clinical dental research, each patient in the data set contains multiple, potentially correlated observations (13). For example, the variable of interest is the time to failure of a restoration placed in a carious lesion and the unit of measurement is the tooth or tooth surface (14). For this reason, when analyzing dental studies, it would be more appropriate to consider the correlation arising from the same individuals and analyze accordingly.

The aim of this study is to choose the most appropriate statistical methods for modeling the explanatory variables of the effect of the time required for tooth decay in individuals with dental plaque. Here, the time required for the formation of caries in the teeth shows the survival time, and each individual shows the cluster effect.

## Materials and Methods

Apart from the classical survival method, there are two approaches in modeling clustered survival data, depending on the purpose of the study. The first of these, conditional models, is based on the principle that the model is randomly modeled and the conditional unobservable effects are valid for each individual. In the second approach, marginal models, and covariates are determined unconditionally. The regression model is assumed to be marginal for each individual, but individuals within groups are associated (15). Conditional models induce dependence by incorporating the random effects, while marginal models directly identify the fixed effects (16). It also keeps some unobserved effect conditional on conditional models, which is modeled randomly for each individual (15). Marginal modeling is an approach that models the marginal distribution of each event time (5). In marginal models, the estimation of β is based on the marginal density of lifetimes. Standard errors are corrected for correlations resulting from clustering in the data. In the parametric case, this probability depends on the marginal distributions. In the semi-parametric case, the partial probability of the ordinary Cox regression model is used. This gives consistent estimates of β in both the parametric and semi-parametric cases (17). In conditional models, cluster effects are considered in the hazard function and used to compare the survival risk of individuals within the same cluster. In marginal approaches, cluster effects are not taken into account in survival risk. It represents the average hazard in the population and is used to compare the survival risk of individuals in the population (18). The marginal model simultaneously models each marginal distribution; however, it also takes into account the correlation structure to determine valid standard errors. For this reason, this approach is considered a design-based model (DBM). The second type of approach, which is the model-based approach (MBM) and called the random effects model, calculates the linear estimator for each cluster based on the random effect, under the assumption that the response variable has a multivariate normal distribution (19, 20).

**Data:** Depending on the purpose of the study, the data to be used in the modeling of the clustered data were simulated. Cox model, frailty model, stratified Cox model, marginal Cox model and marginal weibull models were used to model the generated data. Among these models, frailty model and stratified Cox model are conditional models. The marginal Cox model and the marginal weibull model are marginal models. In the study, simulations consist of three parts. Firstly, the number of units varied as 50, 100, 200, 500, and 1000, assuming that 4 teeth of plaque were taken from the mouth of each individual. In the second stage, the data were handled in an unbalanced cluster size (maximum cluster size 4) and the number of individuals was 50, 100, 200, 500 and 1000. The data produced for the first two conditions were produced with cencor rates of approximately 35%. In the third part, there are cases where the cluster size and number of individuals do not change, but cencor rates change. Simulation scenario; hypothetically generated variables were designed to include bi-class cigarette consumption (smoking status=1, non-smoker status=0), gender (male=1, female=0) and age. Survival time was produced with the gamma distribution. While producing the variables, they were produced in certain proportions in parallel with the purpose of the study. The gender variable was produced as 50% female and 50% male, and the smoking rate was approximately 25% smoking and 75% non-smoking. The age variable obtained from the simulation results varies between 18 and 86. Analyzes were performed with the R programming language (21).

**Model Selection:** There are many criteria for choosing the most appropriate model in statistical analysis. The most commonly used are information and probability-based criteria. Information-based criteria are applied to compare the different models used in the study and the corresponding baseline hazard functions. The most used model selection criteria are Akaike information criterion (AIC) and Bayesian information criterion (BIC) (22-24). In addition, another model selection criterion is corrected Akaike information criterion (AICc). The simulation studies show that AICc tends to overfit as the sample size increases (25). The smallest AIC, AICc and BIC values show that the model is more suitable than other models (26).

## Results

Table 1 gives AIC, AICc, and BIC values at varying sample sizes for each of the five models. When the table is examined, in every case, the best model according to AIC and ACIc is the frailty model. According to BIC, the best model is the marginal Cox model. For all cases, the AIC, AICc and BIC values obtained from the frailty and marginal Cox models were very close to each other. This situation does not change depending on the sample size. However, the worst performing models vary depending on the sample size. The worst performing models in small sample sizes are the marginal Weibull, classic Cox, and stratified Cox models based on all three information criteria, respectively. For example, for the case of k = 4, n = 200, the worst models are the classical Cox, marginal Weibull and strafied Cox models. When the sample size increased further, the ranking became classic Cox, stratified, and marginal Weibull models. In the study, while the best model does not change for all three information criteria, the order of the models with the worst performance changes. The classic Cox model has the worst performance among the models, especially when the sample size increases. According to Table 1, while the worst model in small sample size is marginal Weibull, with the increase in sample size the worst model is obtained as the classical Cox model.

**Table 1.** AIC, AICc and BIC values obtained for the methods when the cluster size is 4 and the number of individuals is 50, 100, 200, 500 and 1000

| **k=4 n=50** | **AIC** | **AICc** | **BIC** |
|---|---|---|---|
| Frailty Model | **501.29*** | **473.20*** | 536.30 |
| Stratified Cox Model | 1059.99 | 1060.05 | 1065.71 |
| Marjinal Cox Modeli | 519.84 | 519.96 | **524.18*** |
| Marginal Wiebull Model | 1537.53 | 1537.73 | 1550.72 |
| Cox Model | 1235.99 | 1236.06 | 1241.72 |
| *k=4 n=100* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **1016.94*** | **943.62*** | 1122.80 |
| Stratified Cox Model | 2451.35 | 2451.38 | 2458.49 |
| Marginal Cox Model | 1086.19 | 1086.26 | **1091.67*** |
| Marginal Wiebull Model | 2996.88 | 2996.98 | 3012.84 |
| Cox Model | 2827.21 | 2827.24 | 2834.34 |
| *k=4 n=200* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **2455.31*** | **2297.66*** | 2736.73 |
| Stratified Cox Model | 5593.87 | 5593.89 | 5602.35 |
| Marginal Cox Model | 2650.35 | 2650.40 | **2657.31*** |
| Marginal Wiebull Model | 5983.01 | 5981.04 | 6001.75 |
| Cox Model | 6315.46 | 6315.48 | 6323.94 |
| *k=4 n=500* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **8264.85*** | **7920.77*** | 9050.86 |
| Stratified Cox Model | 16396.71 | 16396.71 | 16407.03 |
| Marginal Cox Model | 8540.49 | 8540.50 | **8549.52*** |
| Marginal Wiebull Model | 14978.59 | 14978.61 | 15001.00 |
| Cox Model | 18190.29 | 18190.29 | 18200.61 |
| *k=4 n=1000* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **16891.75*** | **16133.08*** | 18847.76 |
| Stratified Cox Model | 36658.51 | 36658.51 | 36670.24 |
| Marginal Cox Model | 17631.59 | 17631.59 | **17641.84*** |
| Marginal Wiebull Model | 30009.91 | 30009.92 | 30035.09 |
| Cox Model | 40275.76 | 40275.76 | 40287.48 |

* Minimum performance criterion value

**Table 2.** AIC, AICc and BIC values obtained for the methods when 75% of the data has a cluster size of 4, 25% of the data has a cluster size of 2 and the number of individuals are 50, 100, 200, 500 and 1000.

| **k=2,4 n=50** | **AIC** | **AICc** | **BIC** |
|---|---|---|---|
| Frailty Model | **379.73*** | **358.58*** | 404.50 |
| Stratified Cox Model | 744.85 | 744.97 | 750.02 |
| Marjinal Cox Modeli | 392.86 | 393.11 | **396.77*** |
| Marginal Wiebull Model | 1148.50 | 1146.66 | 1160.54 |
| Cox Model | 880.54 | 880.67 | 885.71 |
| *k=2,4 n=100* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **854.39*** | **826.36*** | 896.11 |
| Stratified Cox Model | 1720.03 | 1720.09 | 1726.58 |
| Marginal Cox Model | 865.14 | 865.27 | **870.33*** |
| Marginal Wiebull Model | 2225.24 | 2223.32 | 2240.05 |
| Cox Model | 1989.19 | 1989.25 | 1995.74 |
| *k=2,4 n=200* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **2014.57*** | **1944.66*** | 2136.37 |
| Stratified Cox Model | 4020.80 | 4020.83 | 4028.74 |
| Marginal Cox Model | 2043.90 | 2043.96 | **2050.48*** |
| Marginal Wiebull Model | 4562.99 | 4561.04 | 4580.58 |
| Cox Model | 4561.68 | 4561.71 | 4569.61 |
| *k=2,4 n=500* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **5482.83*** | **5251.25*** | 5969.76 |
| Stratified Cox Model | 11762.77 | 11762.78 | 11772.52 |
| Marginal Cox Model | 5621.81 | 5621.84 | **5630.08*** |
| Marginal Wiebull Model | 11197.11 | 11195.13 | 11218.36 |
| Cox Model | 13106.68 | 13106.69 | 13116.43 |
| *k=2,4 n=1000* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **12449.92*** | **12035.41*** | 13465.30 |
| Stratified Cox Model | 26217.62 | 26217.62 | 26228.77 |
| Marginal Cox Model | 12671.78 | 12671.80 | **12681.49*** |
| Marginal Wiebull Model | 22504.32 | 22502.33 | 22528.35 |
| Cox Model | 28921.91 | 28921.92 | 28933.06 |

* Minimum performance criterion value

As a result of the findings, the marginal Cox model with the lowest BIC = 524.18 value was obtained as the best model in the case of k=4, n=50. This condition does not change with the increase in the sample size. For the case of k=4, n=1000, the best model was obtained as the marginal Cox model with the lowest BIC=17641,84 value. Results are similar for AIC and AICc. Regardless of the sample size, the best models according to these criteria are the frailty models. For k=4, n=50 condition, AIC=501.29 and AICc=473.20 were obtained and according to these results, the best model was the frailty model. The results are similar for the case k=4, n=1000 (AIC=16891.75, AICc=16133.08). Considering AIC, AICc and BIC the worst models vary depending on the increase in sample size. For k=4, n=100, the worst model was the marginal Weibull model for all three criteria (BIC=3012.84, AIC=2996.88, AICc=2996.98). The second worst model is the Cox model (BIC=2834.34, AIC=2827.21, AICc=2827.24). In small samples, the worst model is the marginal Weibull model with the highest value among all three information criteria (k=4, n=50; BIC=1550.72, AIC=1537.53, AICc=1537.73). As the sample size increases, the Cox model tends to get the highest values (k=4, n=200; BIC=6323.94, AIC=6315.46, AICc=6315.48). Considering the results obtained, in parallel with the increase in the sample size, the marginal Weibull with the worst performance was obtained as the third best model in the k=4, n=500 scenario. In addition, while the stratified Cox model was the third best model, it was obtained as the second worst model due to the increase in the number of units (BIC=16407.03, AIC=16396.71, AICc=16396.71).

**Table 3**. AIC, AICc and BIC values obtained from the models when k=4, n=500 and censor rates are 10%-70%

| **k=4 n=500 cencor rate %10** | **AIC** | **AICc** | **BIC** |
|---|---|---|---|
| Frailty Model | **10452.58*** | **10071.09*** | 11377.39 |
| Stratified Cox Model | 21983.04 | 21983.05 | 21994.03 |
| Marjinal Cox Model | 10973.92 | 10973.93 | **10983.51*** |
| Marginal Wiebull Model | 21930.08 | 21928.10 | 21952.49 |
| Cox Model | 24473.76 | 24473.77 | 24484.75 |
| *k=4 n=500 cencor rate %20* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **9266.54*** | **8877.32*** | 10182.64 |
| Stratified Cox Model | 19935.11 | 19935.11 | 19945.86 |
| Marginal Cox Model | 9778.20 | 9778.21 | **9787.52*** |
| Marginal Wiebull Model | 18976.66 | 18974.68 | 18999.07 |
| Cox Model | 22146.03 | 22146.04 | 22156.78 |
| *k=4 n=500 cencor rate %30* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **8627.69*** | **8268.72*** | 9456.01 |
| Stratified Cox Model | 17627.24 | 17627.25 | 17637.72 |
| Marginal Cox Model | 8963.91 | 8963.93 | **8973.04*** |
| Marginal Wiebull Model | 16406.22 | 16404.23 | 16428.62 |
| Cox Model | 19558.24 | 19558.25 | 19568.72 |
| *k=4 n=500 cencor rate %40* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **6880.87*** | **6524.40*** | 7660.30 |
| Stratified Cox Model | 15367.90 | 15367.91 | 15378.09 |
| Marginal Cox Model | 7152.82 | 7152.84 | **7161.47*** |
| Marginal Wiebull Model | 13861.21 | 13859.22 | 13883.61 |
| Cox Model | 17043.76 | 17043.77 | 17053.95 |
| *k=4 n=500 cencor rate %50* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **5797.134*** | **5444.264*** | 6538.94 |
| Stratified Cox Model | 12839.01 | 12839.02 | 12848.84 |
| Marginal Cox Model | 6038.85 | 6038.88 | **6047.17*** |
| Marginal Wiebull Model | 11449.82 | 11447.84 | 11472.23 |
| Cox Model | 14233.77 | 14233.78 | 14243.59 |
| *k=4 n=500 cencor rate %60* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **4447.57*** | **4160.64*** | 5011.30 |
| Stratified Cox Model | 10196.36 | 10196.37 | 10205.73 |
| Marginal Cox Model | 4551.78 | 4551.82 | **4559.53*** |
| Marginal Wiebull Model | 8913.29 | 8911.30 | 8935.69 |
| Cox Model | 11312.81 | 11312.82 | 11322.18 |
| *k=4 n=500 cencor rate %70* | ***AIC*** | ***AICc*** | ***BIC*** |
| Frailty Model | **3506.98*** | **3250.38*** | 3982.96 |
| Stratified Cox Model | 7595.61 | 7595.63 | 7604.43 |
| Marginal Cox Model | 3572.68 | 3572.72 | 3579.99 |
| Marginal Wiebull Model | 6836.13 | 6834.15 | 6858.54 |
| Cox Model | 8436.24 | 8436.26 | 8445.06 |

* Minimum performance criterion value

In Table 2, cases where the cluster size is not equal, 75% of the data has a cluster size of 4 and 25% of the data has a cluster size of 2 are represented. The number of individuals varies as 50, 100, 200, 500 and 1000. The aim here is to compare performances between models when cluster sizes are unbalanced. The findings obtained in this context are given in Table 2. For the unbalanced cluster size n=50, 100 and 200 individuals, the best model for AIC and AICc was obtained as the frailty model. The best model obtained for BIC is the marginal Cox model. For all three information criteria, the values obtained from frailty and marginal Cox were close to each other. In addition, according to all three information criteria, stratified Cox model was ranked third, the Cox model was ranked fourth and the marginal Weibull model was ranked fifth. The best models did not change according to the information criteria at n=500 and 1000 individuals. However, as the sample size increased, the worst third model became the marginal Weibull model, the fourth model became the stratified model and the fifth worst model became the Cox model.

Table 3 aims to examine whether there is a difference between model performances at varying cencor rates when k = 4 and n = 500. In simulation studies conducted for this purpose, cencor rates vary between 10% and 70%. When the results for each scenario in the study were examined, the best model was obtained as marginal Cox at each censor rate for the BIC value. The model with the best performance for AIC and AICc values was the frailty model. For each information criteria, the values obtained from the frailty model and marginal Cox model were close to each other. Additionally, marginal Weibull became the third, and stratified Cox became the fourth model. The worst performing model for the entire conditions was the Cox model, which ignores the cluster structure.

According to Table 1, the best models obtained are the frailty model according to AIC and AICc. According to BIC, the best model was obtained as marginal Cox. The ranking of the worst models obtained changes with the increase in sample size. In Table 2, the best models obtained with unbalanced cluster size are found to be similar to Table 1. Again, as in Table 1, the worst models obtained change depending on the sample size. In Table 3, the best models obtained with changing cencor rates are again similar to Tables 1 and 2. In Table 3, the worst models obtained are the marginal Weibull third and the stratified Cox fourth models. The worst performing model for all cases is the Cox model, which ignores the clustering structure.

## Discussion

Clustered data occurs when individuals are divided into different groups and multiple are found in at least a few groups (27). Clustered data are found in many different types of analysis due to the nature of the observations, study design, or the way the data are sampled (28). One of these analyzes is survival analysis. Survival analysis is a method of statistical analysis that deals with events such as the time between a cancer patient's surgery to death, wedding to divorce, and the time between first and second suicide attempts (29). Most epidemiological studies involve survival times clustered into groups, such as families or schools, where some unobserved characteristic possessed by individuals from the same cluster, such as genetic information or unmeasured common environmental exposures, may affect the time to the event under study (30). Survival data are often clustered into groups such as couples, siblings, families, communities, and geographic regions. Oftentimes, observations from the same cluster are tend to be correlated because they share certain characteristics. Ignoring this correlation may lead to biased estimates. That is, if observations from the same cluster tend to be positively correlated, the precision of the predictions is likely to be overestimated (8). There are methods to take into account the dependency caused by cluster effects in clustered survival data. One of them is conditional and the other is marginal model. In conditional models, cluster effects are taken into account in the hazard function and used to compare the survival risk of individuals within the same cluster. In marginal approaches, cluster effects are not taken into account in survival risk. It represents the average hazard in the population and is used to compare the survival risk of individuals in the population (18). AIC, AICc ve BIC were used to compare performances between models.

In this study, it was aimed to obtain the best model by modeling the explanatory variables of the effect of the time required for the decay of the teeth taken from the mouth of each individual for each situation. Within the scope of the findings, the best model according to AIC and AICc is the frailty model. According to BIC, it is the marginal Cox model. In addition, when all the results are examined, the, AIC, AICc and BIC values obtained for the frailty model and the marginal Cox models were very close to each other. As the sample size increases, the classical Cox model becomes the model with the worst performance.

In parallel with our study, Yashin and Iachine showed how a bivariate survival model based on the concept of correlated individual frailty can be used for genetic analysis times. Six genetic frailty models were used in the study and applied to twin survival data in Denmark (31). Tessema et al. used regional status of women in Ethiopia as cluster effect. They aimed to model the determinants of time to first marriage with frailty models and used AIC for model performances (32). Fagbamigbe et al compared 13 survival regression models to assess factors associated with the timing of reconstruction complications in implanted teeth in a Swedish cohort. They used AIC and BIC to determine which model was better (33). Lorino et al. suggested in their study that the classical Cox model gives biased results when survival data are clustered in small groups, as is the case with many data types. In their study, they proposed two different models for multivariate survival data: frailty and marginal model. As obtained from the results of the study, they stated that it would be more accurate to use models that take correlation into account in large samples rather than small samples (34). Mahmood et al. stated that observations taken from the same cluster are correlated and therefore share unobserved characteristics among individuals, and ignoring the correlation between observations leads to inaccurate estimates of standard errors of the variables of interest. For these reasons, they suggested that frailty models should be preferred instead of the classical Cox model (35). Thapa et al. compared the model performances by applying the Cox model and frailty models to the application data in their study. According to their results, they found that the frailty model had a smaller AIC value and higher predictive ability. In parallel with the study, the frailty model has the smallest AIC value in that study (36). In parallel with the findings obtained from the study, as the sample size increases, the importance of using models that take into account the effect of dependency becomes evident.

There are some limitations in this study. In the study, simulations were made only for some scenarios. Different simulation scenarios can be tried; the number of explanatory variables used to compare model performances can be increased. Additionally, the study can be supported by applying it on a real data set. It is recommended that these situations be taken into consideration in future studies.

In conclusion of the study, ignoring the data type in data with a clustered data structure, as in dental data, may lead to biased estimations. In this type of data, it is recommended to use the frailty or marginal Cox model, taking into account the performance criterion. In cases where the data structure is clustered in new health studies to be conducted, it is recommended to use appropriate methods that take this structure into account.

### References

1. Leung KM, Elashoff RM, Afifi AA. Censoring issues in survival analysis. Annu. Rev. Public Health 1997; 18(1): 83-104.
2. Moore DF. Applied Survival Analysis Using R. 5th edition Switzerland: Springer, 2016.
3. Collett D. Modelling Survival Data in Medical Research. 3th edition London: Chapman and Hall/CRC, 2015.
4. Guo, S. Survival Analysis. 1th Edition, New York: Oxford University Press, 2010.
5. Hou B. Sequential Analysis of Clustered Survival Data by Marginal Methods. Graduate Program in Statistics. Doctoral Thesis, New Jersey: The State University of New Jersey, 2008.
6. Gangnon RE, Kosorok MR. Sample-size formula for clustered survival data using weighted log-rank statistics. Biometrika 2004; 91(2): 263-275.
7. Gregg ME, Datta S, Lorenz D. A log rank test for clustered data with informative within-cluster group size. Statistics in Medicine 2018; 37(27): 4071-4082.
8. Guo G, Rodriguez G. Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala. Journal of the American Statistical Association. 1992; 87(420): 969-976.
9. Gönen M, Panageas KS, Larson SM. Statistical issues in analysis of diagnostic imaging experiments with multiple observations per patient. Radiology 2001; 221(3): 763-767.
10. Taboada-Aranza O, Rodríguez-Nieto K. Prevalence of plaque and dental decay in the first permanent molar in a school population of south Mexico City. Bol Med Hosp Infant Mex. 2018; 75(2): 113-118.
11. Addy M. Plaque control as a scientific basis for the prevention of dental caries. Journal of the Royal Society of Medicine 1986; 79(Suppl 14): 6.
12. Lee HJ, Kim JB, Jin BH, Paik DI, Bae KH. Risk factors for dental caries in childhood: a five-year survival analysis. Community Dentistry and Oral Epidemiology 2015; 43(2): 163-171.
13. Chuang S, Tian L, Wei L, Dodson T. Predicting dental implant survival by use of the marginal approach of the semi-parametric survival methods for clustered observations. J Dent Res 2002; 81(12): 851-855.
14. Wong MC, Lam K, Lo EC. Multilevel modelling of clustered grouped survival data using Cox regression model: An application to ART dental restorations. Statistics in Medicine 2006; 25(3): 447-457.
15. Eriksson F. Correlated Random Effects Models for Clustered Survival Data. Department of Mathematical Sciences. Doctoral Thesis, Goteborg: Chalmers University of Technology and University of Gothenburg, 2012.
16. Segal MR, Neuhaus JM, James IR. Dependence estimation for marginal models of multivariate survival data. Lifetime Data Anal 1997; 3: 251-268.
17. Z Zetterqvist J. Proportional Hazards Model for Matched Failure Time Data. Department of Mathematics. Master's Thesis, Stockholm: Stockholm University, 2013.
18. Gorfine M, De-Picciotto R, Hsu L. Conditional and marginal estimates in case-control family data–extensions and sensitivity analyses. J Stat Comput Sim. 2012; 82(10): 1449-1470.
19. Brown H, Prescott R. Applied mixed models in medicine. 3th Edition, United Kingdom: John Wiley & Sons, 2014.
20. Dağoğlu Hark B. Bilgilendirici Küme Boyutuna Sahip Kümelenmiş Verilerin Marjinal Modellemesi İçin Yeni Bir Yaklaşım. Doctoral Thesis, Adana: Çukurova University, 2020.
21. Akbaş KE. Kümelenmiş Verilerde Sağkalım Analizi. Doctoral Thesis, Malatya: Inonu University 2022.
22. Brettschneider J, Burgess M. Using a frailty model to measure the effect of covariates on the disposition effect. CRISM Working Paper Series, 2017; 17(5):1-46.
23. Xu R, O'quigley J. AR2 type measure of dependence for proportional hazards models. Journal of Nonparametric Statistics 1999; 12(1): 83-107.
24. Kiche J, Ngesa O, Orwa G. Misspecification Of Frailty Random Effects In A Clustered Survival Data. Journal of Statistical and Econometric Methods 2019; 8(2): 1-25.
25. McQuarrie A, Shumway R, Tsai C-L. The model selection criterion AICu. Statistics & probability letters. 1997; 34(3): 285-292.
26. Wanigasekara C, Nawarathna LS, Vithanaarachchi V. Frailty Models for Predicting Eruption Time and Sequence of Permanent Dentition in Sri Lankan Children. Asian J Probab Stat 2021; 11(3): 1-10.

27. Holodinsky JK, Austin PC, Williamson TS. An introduction to clustered data and multilevel analyses. Family Practice 2020; 37(5): 719-22.

28. Graubard BI, Korn EL. Regression analysis with clustered data. Statistics in Medicine. 1994; 13(5-7): 509-522.

29. Knox KL, Bajorska A, Feng C, et al. Survival analysis for observational and clustered data: An application for assessing individual and environmental risk factors for suicide. Shanghai Arch. Psychiatry 2013; 25(3): 183.

30. Gorfine M, Zucker DM, Hsu L. Prospective survival analysis with a general semiparametric shared frailty model: A pseudo full likelihood approach. Biometrika 2006; 93(3): 735-41.

31. Yashin AI, Iachine IA. Genetic analysis of durations: correlated frailty model applied to survival of Danish twins. Genet. Epidemiol 1995; 12(5): 529-38.

32. Tessema B, Ayalew S, Mohammed K. Modeling the determinants of time-to-age at first marriage in ethiopian women: A comparison of various parametric shared frailty models. Sci J Public Health 2015; 3(5): 707-18.

33. Fagbamigbe AF, Karlsson K, Derks J, Petzold M. Performance evaluation of survival regression models in analysing Swedish dental implant complication data with frailty. Plos One 2021; 16(1): 1-16.

34. Lorino T, Sanaa M, Robin S, Daudin JJ. Comparison of semiparametric regression models for correlated survival data using simulations. Communications in Statistics-Theory and Methods 2004; 33(8): 1975-91.

35. Mahmood S, Zainab B, Latif AM. Frailty modeling for clustered survival data: an application to birth interval in Bangladesh. J Appl Stat 2013; 40(12): 2670-80.

36. Thapa R, Burkhart HE, Li J, Hong Y. Modeling clustered survival times of loblolly pine with time-dependent covariates and shared frailties. Journal of Agricultural, Biological, and environmental Statistics 2016; 21: 92-110.