Analysis of Birth Spacing Using Frailty Models

Arezou Bagheri¹, Mahsa Saadati²

1. National Population Studies & Comprehensive Management Institute, Tehran, Iran
2. National Population Studies & Comprehensive Management Institute, Tehran, Iran

Abstract

Background and objectives: Birth spacing is an important variable for identification of fertility acceleration, total fertility rate, and maternal and fetal health. Therefore, special attention has been paid to this issue by researchers in the fields of medical sciences, health, and population. In addition, proper analysis of this concept is of foremost importance. Application of classical analytical techniques with no attention to their assumptions (e.g., independence of events) is associated with inefficient results. As such, this study aimed to present frailty models as effective models for this analysis.

Methods: Frailty models consider the dependence between unobserved intervals and dispersions by exerting a random impact on the model. Different types of these models include shared, conditional, correlated and time-dependent frailty, each of which along with their applications were presented in the current research using two examples.

Results: In practice, the shared frailty model is highly applied due to its simplicity. Nevertheless, since most of the unknown factors affecting the birth spacing are not common between different births, the shared frailty models must be used with caution.

Conclusion: Use of classical statistical methods, such as the Cox proportional hazards model, the important assumption of which is the dependence of events occurred, is not appropriate for the accurate analysis of birth spacing. On the other hand, frailty models consider the correlation between the intervals and are an effective method for analysis of birth spacing, use of which is recommended to researchers in fields of medicine and population.

Keywords: Fertility, Birth Spacing, Frailty Model
result, a high fertility rate, especially when the first child is a girl (2-4). Short spacing not only leads to increased population growth but also negatively affects maternal and fetal health (5). In addition, consecutive births reduce maternal strength and often lead to preterm delivery, anemia, bleeding in late pregnancy, stillbirth, and miscarriage (5, 6). Furthermore, several studies have demonstrated that shorter birth spacing, especially under two years, is closely associated with neonatal health risks, including early birth and low birth weight.

The World Health Organization (WHO) recommends a minimum of two years between births to reduce neonatal mortality and improve maternal health. On the other hand, the United States Agency for International Development (USAID) considers an optimal birth spacing to be three-five years. According to the literature, if births occur within a 36-month interval in less-developed countries, the neonatal and child mortality rates will increase by 2% and 35%, respectively (1).

Birth spacing has been one of the most important strategies for improving maternal and neonatal health over the past two decades in Iran, and several studies have been conducted in this area recently (6-9). Factors such as the duration of breastfeeding (5, 10), the survival and gender of the previous child (10, 11), maternal level of education (11, 10), age of women at the time of delivery (5, 10, 12), the interval between the last two childbirths (5), and the socioeconomic status of women (12), can affect birth spacing. Since these factors vary among different populations, their analysis is significantly vital (1).

Common statistical methods for analyzing birth spacing, irrespective of their occurrence, are models such as logistic regression that allow researchers to investigate the relationship between the probability of childbearing and the predictor variables of covariates studied. Nevertheless, these methods ignore the knowledge of childbearing time and only evaluate the probability of occurrence, failing to find an answer to the question of “when does the childbirth happen?"

The appropriate method for determining birth spacing is survival analysis (13), which is a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. One of the most important features of survival data is that the occurrence time of the event is invisible or censored, generally identified as not knowing the exact survival time. Therefore, the exact time of an event is known to those who have access to the data (uncensored data) and unknown to those who have no access to the data (censored data).

Conventional methods are not ideal techniques for dealing with censored data. Another reason for the lack of possibility of using methods such as simple linear regression in the analysis of survival data is their asymmetrical distribution since the survival time is often positive for a group of people with skewness. Therefore, it is not logical to assume that these data have a normal distribution (14). The Cox proportional hazards model is the most conventional survival analysis method for evaluating the effect of predictor variables of the incident time without the need to determine the basic risk function.

A basic assumption for the validity of the Cox model is the proportionality of the hazards or the independence of the event times of each other, which is often ignored in
applications of this model. Meanwhile, in most studies, including those on the assessment of birth spacing, there is a correlation between the occurrence times of events (childbearing times). In these studies, using the Cox model while ignoring the correlations between the intervals leads to an error in estimating the standard deviation of the desired parameters and results in wrong inferences (14, 15). Therefore, various statistical methods have been presented from generalizing the Cox model for the correct analysis of this type of data in biostatistics, epidemiology and medicine.

The variance-corrected and frailty models are some models obtained from the generalization of the Cox model and have been used for assessing the correlated survival data in the past years. In the variance-corrected models (conditional and marginal models), correlation is considered in the estimation of the variance of the parameters in the model and correct inferences for data analysis are obtained. Nonetheless, these models ignore internal correlation and distribution among different people. In conditional variance-corrected models, it is only possible to compare the time of an event between different individuals (and not just the internal correlation of a person).

Frailty models that consider the correlation between events by exerting a random impact on the model are more efficient in entering dispersion into the model, compared to the variance-corrected models. In addition, frailty models are useful when the level of dependency between the occurrence of events for an individual and their correlation descriptions is considered. In these models, we can assess the occurrence of an event for a person and assess the way correlation changes between individuals. Moreover, these models provide information on the relationships between the occurrence times of events (16, 17).

Considering the importance of studying the birth spacing, using valid statistical methods in this area is essential. With this background in mind, this study aimed to present frailty models for analysis of these data. In section (2), different types of frailty models, including shared, conditional, correlated and time-dependent models, are presented. Application of these models in the analysis of birth spacing is explained in section (3) with two examples. Finally, the issue is discussed and concluded at the end of the study.

**Materials and Methods**

**Materials and Methods: Frailty Models**

When several events occur for each person during the period of follow-up, individual factors create a correlation between the occurrence times of events for the person, which can be the reason for differences in individuals (dispersion among individuals). Some of these factors are known and are measured with predictor variables. However, it is clear that in various studies, all factors are not collected for reasons such as financial constraints, immeasurability, or unknown nature.

Conventional models of survival analysis (e.g., the Cox model) lead to overestimated model parameters by ignoring individual effects. In recent studies, a random component is used to express these unknown factors and to create a correlation between the occurrence of an individual’s events. This random component that acts as a random effect in longitudinal models is called frailty in survival analysis. In the simplest form, frailty is an unobserved accidental agent that corrects the risk function associated with a person or related individuals. In addition, frailty is a statistical concept that considers
the dispersion of unmeasured predictor variables in the model (18).

With the presence of frailty in the model, the dispersion of occurrence time of events is considered in two different sources of risk and frailty function (19). In the data of correlated events, dispersion between individuals and the correlation between events establish two correlation sources (17). The frailty in the model implies both sources of correlation, which are not expressed through the predictor variables, in the model. The risk function of this model for the \( j \)th event of the \( i \)th person is expressed in the form of equation 1:

\[
\lambda_{ij}(t) = \lambda_0(t) \omega_i \exp(\theta^T Z)
\]

In this model, \( \omega_{ij} \) is frailty, which has a multiplicative effect on the hazard function. Therefore, the large amount of frailty increases the risk of occurrence of an event for a specific person. In addition, \( \lambda_0(t) \) is the shared basic risk function, \( Z \) is the vector of predictor variables and \( \beta \) is the vector of model’s coefficients. In this model, the variance of frailty determines the degree of relationships between the time of events.

Various frailty models, including shared, conventional, correlation and time-dependent models, are present to fit the data, which are assessed in this part of the research.

1. Shared Frailty Model

In this model, a frailty is considered for each person to link the dispersion of risk function from a person to another individual and the correlation of internal personal events to unseen individual characteristics (19). This risk function of this model for the \( j \)th event of the \( i \)th individual is expressed in the form of equation 2:

\[
\lambda_{ij}(t) = \lambda_0(t) \omega_i \exp(\theta^T Z)
\]

It is assumed that the events are independent of each other in case of frailty. In other words, frailties create the correlation between the occurrence times of events. This leads to the normal performing of the deductions, and the observations become dependent without being conditional. The word shared means that an equal individual effect exists for occurrence times of events for a person that remains constant over time (16, 17). By entering frailty into the model, self-frailty is equal to one if its variance were zero (since it is assumed in most cases that the mean frailty is equal to one), thereby obtaining a non-frailty model. If frailty exists (i.e. the variance of frailty is more than zero) and its value is more than one for a person, it is indicative of a higher possibility of incidence of the event for the person. It means that the person experiences a higher number of events per unit time. However, the interpretation will be reverse if frailty is below one (20).

Since the coefficients of the predictor variables in this model are estimated at the individual level and are estimated on the condition of unobserved frailty, their interpretation is not simple and should be done with caution. However, they can be interpreted as the effect of a risk factor for a particular person. Shared frailty models have some limitations; for instance, frailty only expresses the part of unknown factors that exists between shared events while these factors are not necessarily equal in practice. In addition, while only the positive relationship between events of a person is considered by shared frailty, this relationship can be negative in some conditions.

On the other hand, given the fact that the dependence of the occurrence time of events within a person is determined on the basis of...
marginal distributions and in a shared form, the dependence and dispersion of the society are not separable and are expressed in a damaged and merged form (19).

2. Conditional Frailty Model

In shared frailty models, such as the Cox model, the basic risk function does not change between events and dispersion is considered directly through the random effect in the model. With the constant basic risk of events, the effects of predictor variables cannot be estimated separately. The best way to consider dispersion between individuals and the dependence between events is to use frailty and classification of the basic risk function simultaneously in the model, which is realized in the conditional frailty model (19). The conditional frailty model is a model in which the random effect for expressing dispersion between individuals is combined with the classified basic risk for each event. \( \lambda_{ij} \), the risk of the occurrence of the \( j \)th event for the \( i \)th person, is expressed for the time interval between events, as follows:

\[
\lambda_{ij}(t) = \lambda_0(t-t_{i-1}) e^{\omega_{ij}}
\]

Where \( \lambda_0 \) is the basic risk that changes with the number of the event and \( (t-t_{i-1}) \) is the time interval. In this model, the \( i \)th person has a shared random effect, which is constant over time.

3. Correlated Frailty Model

In the shared frailty model, there is only one unobserved shared factor for all events that occur for a person, which leads to correlation and dispersion. Meanwhile, it is possible that each person or event is an unobserved factor that is different from others. Therefore, assuming that frailty is shared creates restrictions on the model. To prevent this issue, a model with multivariate and correlated frailty is recommended. In the correlated frailty model, there are more parameters in addition to frailty variance that can express the correlation between frailties in each person (group or cluster).

Given the fact that dispersion is introduced into this model, it is sometimes observed that the standard error of the fixed effects is slightly higher than the shared frailty model and the confidence interval becomes larger (20). The risk function in the correlated frailty model is always similar to equation (1), in which \( \omega_{ij} \) shows correlated frailty and is different from an event to another event. It should be noted that this model is more applied in multivariate or panel survival models.

4. Time-dependent Frailty Model

In shared frailty models, it is assumed that the frailty of individuals is constant over time, dependent from each other and equally distributed. Nevertheless, these assumptions are not correct in some conditions. Experience has shown that dependencies between different time periods of an event often decrease when they get far from each other in terms of time. Therefore, use of a model with constant frailty is not sufficient for expressing the changeability in this condition since this changeability might not be completely analyzed by frailty or known predictor variables dependent on time. Therefore, it is necessary to evaluate the effect of the occurrence of the pattern through a time-dependent frailty in data of correlated events (10, 20).

In the time-dependent frailty model, it is possible for a person to evaluate changes in the occurrence of an event from the first event over time. If the relationship is positive and significant, it means that the occurrence risk of an event increases with the occurrence of the first event (15). In a model where the response variable is the time interval between
the occurrence of events, the time-dependent frailty model can be time from the first event to the \( j \)th event. Since this model has several complications, it is less applied in practice.

**Results**

**Application of Frailty Models to Assess Birth Spacing**

In this section, two studies conducted in Iran (21) and Bangladesh (22), which used the shared frailty model to evaluate the factors affecting birth spacing, are presented.

1. **Evaluation of Factors Affecting Birth Spacing in Zarrin Dasht, Fars Province, Iran**

   In a cross-sectional research performed in Zarin Dasht in 2014, data related to 1064 women of childbearing age were collected to evaluate the factors affecting birth spacing and determine the optimal birth spacing. The shared frailty model was applied to properly analyze the data by considering the first-sixth deliveries of the subjects. Results of this modeling are presented in Table (1), according to which the variables of marriage age, occupational status, and better socioeconomic status of women accelerated the next pregnancy (the coefficients of variables in the model are positive). While the length of the spacing for the previous child was also significant, by increasing each year, only about 0.3% of the risk of earlier childbirth increases, which may not be that significant. Increased level of education, age at delivery, the length of breastfeeding of the previous child, the difference in the age of the parents, and the employment status of men leads to increased birth spacing and the rate of occurrence of the next childbirth decreases. For instance, with an increase in a woman's level of education, her childbirth rate is reduced by 47% (relative risk=0.53). Despite the significant difference in parental age, only four percent of the risk of next birth is reduced. Meanwhile, this value is about nine percent per increase of each six months of breastfeeding of the previous child.

| Table 1. Frailty Model in Study of Birth Spacing in Zarin Dasht, Fars Province, Iran |
|-----------------------------------|-----------------|-----------------|-----------------|
| **Variable**                      | **Variable Coefficient in the Model** | **Relative Risk** | **P-value**     |
| Marriage age of women (year)      | 1.36            | 3.90            | <0.001          |
| Age difference of parents         | -0.04           | 0.96            | 0.013           |
| Occupational status of women      | 1.33            | 3.77            | 0.002           |
| Level of education of women       | -0.63           | 0.53            | <0.001          |
| Socioeconomic status              | 0.90            | 2.47            | <0.001          |
| Occupational status of men        | -0.76           | 0.47            | 0.004           |
| Previous spacing length (month)   | 0.003           | 1.003           | 0.038           |
| Breastfeeding length of previous child (six months) | -0.01 | 0.98 | <0.001 |
| Women’s age at delivery (year)    | -1.70           | 0.18            | <0.001          |
| Frailty                           | 0.47            |                 | <0.001          |
The significance of frailty in this model demonstrated that the correlation between the birth spacing of each person and the individual difference is an important factor for modeling (21).

2. Evaluation of Factors Affecting Birth Spacing in Bangladesh

In 2016, a research was conducted by Islam on the data of Bangladesh Demographic and Health Survey (BDHS, 2011) to evaluate the factors affecting birth spacing between the first two deliveries, including age at marriage, level of education, occupational status, economic status, province and place of residence of women along with gender and survival rate of the first child. Since the chance of the birth of the next children decreases if the second birth interval were long, only the first two deliveries were evaluated in the mentioned research (22). Table (2) shows significant variables on the birth spacing analyzed using the frailty model. As observed, the frailty coefficient in this model was significant and ignoring it would lead to inappropriate inferences. In addition, results demonstrated that women residing in cities and having more than one live birth had a longer birth spacing, compared to women residing in rural areas, who have a low level of education and a stillbirth as the first child (22).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Variable coefficient in the model</th>
<th>Relative risk</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place of residence</td>
<td>Village (reference variable)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>City</td>
<td>-0.140</td>
<td>0.870</td>
<td>0.002</td>
</tr>
<tr>
<td>Level of education of women</td>
<td>Illiterate (reference variable)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primary education</td>
<td>-0.101</td>
<td>0.904</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>High school diploma</td>
<td>-0.380</td>
<td>0.654</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Academic education</td>
<td>-0.378</td>
<td>0.685</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Survival status of the first child</td>
<td>Stillbirth (reference variable)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alive</td>
<td>-1.026</td>
<td>0.358</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Frailty</td>
<td></td>
<td>0.44</td>
<td></td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Discussion

Birth spacing involves two components of fertility process, including quantum of fertility, which is the ratio of women who pass the process of reaching the next child, and tempo of fertility, which is the time required to transfer to the next stage for continuing fertility (23). Therefore, special attention has been paid to the analysis of this issue by demographers. On the other hand, since this variable plays a role in determining the size of the household and maternal and neonatal mortality rates, it is also considered by doctors and researchers in the area of health sciences. Therefore, proper modeling and its analysis in these fields are of paramount importance.

Today, several studies have been conducted in Iran and the world to assess the factors affecting birth spacing using the Cox model (23-25). In 2002, Xavier & Padmadas reviewed the family health database of India applying the birth table and Cox model to evaluate the relationship between predictor variables and times of birth (26). Using the Cox model, Suwal (2001) evaluated the impact of sociocultural dynamics on birth spacing in Nepal (27). In addition, Ghilagaber and Gyimah (2004) assessed birth spacing to study fertility quantum and tempo using survival parametric models in Ghana (28). In order to evaluate the dynamics and spacing of births in Ethiopia, Beaujot & Sahleyesus (2007) used the log normal accelerated failure time model to assess the relationship between predictor variables and times of birth (29).

In a research by Rasekh and Momtaz (2008), the Cox regression model was exploited and the correlation between the birth times was ignored. According to their results, various factors, including the level of education of women and the use of contraceptives, affected birth spacing (24). On the other hand, Eyni Zeynab & Agha (2005) used the population and health plan data (DHS) in 2000 in Iran and exploited the life table, the Cox model and the Weibull parametric model to assess the socioeconomic factors affecting the interval birth the first two children and birth spacing (30). In a research by Bagheri and Sa’adati in 2016, parametric models were used for modeling of the proper interval between births (31).

In all studies where birth spacing was modeled regardless of considering the correlation between spaces, the analysis units are recorded as birth spacing for each individual, and for Cox or parametric survival models are applied for their analysis.

In the Cox model, it is assumed that the distances are independent of each other and all dispersions can be explained by the selected predictors. Meanwhile, there is a correlation between the birth intervals of a woman, and unobserved dispersions exist in the model due to various reasons, including being unmeasurable or ignoring other predictors (32). If unmeasurable frailties exist, risks will not only be a function of predictors and will be also related to the unmeasurable frailties. Moreover, given the fact that the children of a woman are equal and significantly similar to each other, compared to the time they are randomly selected, it is more accurate to consider women as a cluster along with the correlation among children (23). If unobserved frailty exists in the model, the risk will be affected by frailty in addition to the observed predictor variables.

Various studies have shown that ignoring frailty leads to a decrease or increase of risk with a significantly higher and lower
acceleration, respectively. Therefore, correcting the conventional models by using frailty, which involves dispersion resulting from unmeasurable factors, is a method for modifying changes in determining the model (33). Recently, there has been a rise in the use of correlated survival models for analyzing birth spacing. In 2008, Pushkar & Samistha evaluated the birth spacing, planned or unplanned pregnancy, and survival rate of infants using the correlated risk model (32). In addition, Islam (2013) applied the frailty model for modeling of birth spacing in Bangladesh (22). Similarly, the frailty model was exploited by Zare et al. to assess the birth spacing of women in Zarin Dasht, Iran (21).

Conclusion
Considering the importance of using proper models for analysis of birth spacing, the necessity of entering correlation between intervals and paying attention to the fact that the children of a woman are similar to each other was expressed in the current study. Since frailty models enter these correlations based on a random effect that is able to show the unmeasurable dispersion that is not independently expressed by predictors and provide the ability to have a person-specific interpretation of the estimations (similar to mixed models in analysis of longitudinal data), these models are recommended as a valid method for analysis of these intervals (33).

Acknowledgements
Not applicable

Declarations

Funding source(s)
This article was extracted from the thesis entitled “analysis of birth spacing using event log models” (report No: 20/30022, date: 09/16/2017) and was financially supported by the National Population Studies & Comprehensive Management Institute in 2017.

Conflict of interest
We declare that we have no financial or non-financial conflicts of interest related to the subject matter or materials discussed in the article.
References


