Analyzing Menstruating Interval to the First Conception: An Application of CPH and AFT Models

Brijesh P. Singh[1] and Hricha Rai[1]*

[1] Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India
* Hricha Rai, corresponding author. Email: hricharai@bhu.ac.in
Submitted: 16 August 2024. Accepted: 28 February 2025. Published: 25 April 2025
Volume 34, 2026. pp. 199–216. http://doi.org/10.25133/JPSSv342026.011

Abstract

Survival analysis techniques analyze the time-to-event data. In survival analysis, there are two essential methods: Semiparametric and parametric methods. The semiparametric Cox proportional hazard (CPH) model is frequently used to analyze demographic survival data. However, the presumption of proportional hazard (PH) is not always met in real-life data. Meanwhile, one can use parametric Accelerated Failure Time (AFT) models as an alternative when the PH assumption does not hold. The main purpose of this article is to analyze and compare the performances of the CPH model and AFT models in identifying the significant covariates affecting the menstruating interval to first conception (MIFC). In this article, three parametric AFT distributions based on Exponential, Weibull, and Log-normal distributions are used to check the performance. We have shown the violation of having a proportional hazard assumption with the help of a graphical technique and statistical test. According to the Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC), we have found that the Weibull AFT model is best suited for the data. Further in this paper, we have identified the significant factors affecting the MIFC. Analysis reveals that covariates such as age at marriage, place of residence, wealth index, mother's education, and Mass media exposure are significant factors affecting the menstruating interval to first conception in Uttar Pradesh.

Keywords

AFT models; CPH model; menstruating interval to the first conception; PH assumption; survival analysis; Weibull AFT model

Introduction

Survival analysis is a field of statistics dedicated to the study of time-to-event data. The Cox Proportional Hazard Model (CPHM) is frequently utilized in real-life scenarios involving time-to-event data to analyze the effect of multiple factors on survival time. The Proportional Hazards (PH) model (Cox, 1972) is semiparametric, meaning that the baseline hazard function (hazard function in the absence of all the covariates) is left unidentified, and no assumptions are made about the distributional pattern of survival times. The PH model assumes that the hazard ratios between different levels of the covariates are constant over time. If this assumption is violated, the model may produce bias or misinterpret the estimate (Schemper, 1992).

Another model in survival analysis that also relies on the proportional hazards (PH) assumption is the Parametric PH model (Collett, 2023). These models are the upgradation of the semiparametric PH model by the assumption that the baseline hazard $(h_0(t))$ follows a specific distribution. Although the parametric PH model has numerous applications in clinical and demographic research (Singh et al., 2016; Zhang, 2016), it cannot serve as an alternative to the semiparametric PH model when dealing with non-PH data. In contrast, the AFT model is a parametric approach in survival analysis that can be utilized instead of the PH model in cases of non-proportion hazard in the data (Kay et al., 2002).

Another primary advantage of the AFT model is its ability to handle time-dependent covariates more effectively than the CPH model (Mendes & Fard, 2014). The assumptions underlying the AFT distribution is that the survival and hazard function involved in the model follow some fixed probability distributions, the parameters of which can be estimated. The superiority of the AFT distribution over the CPHM is also in the explanation of the results, which becomes easier in the former case as the parameters in the AFT distribution indicate the effect of the independent variables on the mean survival time instead of the hazard rate (Kalbfleisch & Prentice, 2011).

Many researchers have widely used the parametric AFT model in clinical research. According to Qi (2009), AFT distribution fits better with the HIV data than the Cox PH model. The application and comparison of the CPHM and AFT distributions are done for several censored data using a simulation study (Orbe et al., 2002) and also on real-life data (Bakhshi et al., 2017; Faruk, 2018; Swindell, 2009; Tripathy et al., 2022). The AFT distribution is also recommended by many authors in analyzing the time-to-event data in clinical studies (Khanal et al., 2014). However, the literature on comparative studies between the CPH model and AFT models on birth interval data is limited in India, especially with a large amount of real-life data.

Also, in our research, we are considering the duration between marriage and time to first conception instead of the first birth as no woman will conceive before her marriage in the Indian societal setup because of the cultural norms associated with the marriage, so the interval will start from at least after nine months (the most common gestation period) of marriage for the total considered duration in case of first birth data and the distribution used for such a scenario should be truncated (Sheps et al., 1970). This duration between the date of marriage and the date of the first clinically confirmed conception is called the menstruating interval to the first conception (MIFC).

The rationale for using MIFC as a measure in population studies is grounded in its ability to provide insights into fertility, reproductive health, and population dynamics. MIFC also provides insights into cultural and behavioral patterns in a population, such as the timing of marriage, contraceptive use, and societal norms regarding family planning. Variations in MIFC may reflect changing attitudes toward childbearing. In our study, we have considered only complete conception that results in live births (Wood, 2017).

In a traditional society, where the use of contraception is not very prevalent before the first birth, several unique factors may influence the waiting time distribution for the first conception (Nath et al., 1995; Singh et al., 2016). This study helps in identifying the factors that are responsible for early or delayed conception. In our study, we have used three AFT models having distributions as Exponential AFT, Weibull AFT and log-logistic AFT, and made the comparison among these with the PH model and chose the best with the help of AIC value.

Study objectives

- To check whether the PH assumption is followed by the Menstruating interval for the first conception data.
- To compare the performance of the CPH model and considered AFT models.
- To identify the covariates affecting the MIFC according to the various socio-demographic characteristics of women with the help of the AFT model.

Materials and methods

The data used in this article was obtained from the fifth round of the National Family Health Survey (NFHS–5), which was conducted from 2019 to 2021 by the International Institute for Population Sciences (IIPS) and ICF (2022). This study used information from 57,446 ever-married women aged 15–49 years in Uttar Pradesh. The analyses in this article focus on first births. Since the commitment and risk of having a first baby for a couple are usually settled at first conception, the timing of conception leading to a live birth is used, rather than the timing of the birth of the child itself, as the dependent variable. The survey data provides information on the timing of first births. The time to the occurrence of first conception is calculated by the information on birth data (on average, 9 months preceding childbirth), and a measure of the MIFC is constructed—the dependent variable analyzed in this study.

Previous research has successfully employed this approach (Ghimire & Axinn, 2010). Once a female becomes pregnant, the female is considered as a case in the study as the event of conception has occurred. Females who were not yet pregnant up to the study period were considered censored cases. A woman is regarded as being at risk of complete conception only after her marriage, considering our societal set-up, as in India, cohabitation outside marriage is uncommon, and premarital childbearing is rare (Chandra-Mouli et al., 2013; Talwar, 1967). To avoid any bias, only those females who have a waiting time for their first conception of up to 120 months are included. Out of the total sample, 5,240 cases (almost 9% of the total sample) are censored, i.e., they have not given birth to any live child, and neither are they pregnant till the period of the survey date. All the females included in the study have complete information on the factors affecting the menstruation interval to first conception, which has been considered in the study, such as respondent's age, place of residence, religion, caste, respondent education, wealth index, and age at first marriage.

The tri-mean, spread, and frequency distribution are used as descriptive statistics to show the summary of the sample used in this study. The trimean is often used in statistics to provide a robust measure of central tendency that is less affected by outliers compared to the mean in the case of non-normal data (Páez & Boisjoly, 2022). The tri-mean can be calculated as

$$Tri-mean = \frac{Q_1 + Q_3 + 2Q_2}{4}$$

The spread used in this article is semi-interquartile range (SIQR), which is particularly useful in robust statistics because it is less sensitive to outliers than other measures of spread, such as the standard deviation (Wilcox, 2012). The SIQR is given as

$$SIQR = \frac{Q_3 - Q_1}{2}$$

In this article, survival analysis is done to analyze the factors affecting the duration of MIFC. In the first part of the study, the semiparametric Cox's proportional hazard regression technique (Cox, 1972) is used to analyze the significant factors affecting the duration, and with the help of a graphical method and some statistical tests (Schoenfeld, 1982), the proportionality assumption of the covariates is checked. In case of violation of proportionality assumption, parametric survival modeling, e.g., AFT models (Exponential AFT distribution, Weibull AFT distribution, and Loglogistic AFT distribution), is applied. The best model is chosen using Akaike Information Criteria (AIC) (Akaike, 1974). The best-fitted model has the lowest AIC value.

Cox's Proportional Hazard model

The Cox's PH regression model gives the instantaneous failure rate (conception rate in this study) at time *t* and is given by the following equation.

$$h(t) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_p x_p)$$

Where $h(t)$ is the hazard rate at the time *t*, $h_0(t)$ is the baseline hazard rate (the expected hazard in the absence of any covariate), $\beta_1, \beta_2, \ldots..\beta_p$ are the regression coefficients, and $x_1, x_2, \ldots..x_p$ are *p* independent variables (Socio-demographic variables of respondents in our study). The ratio $h(t)/h_0(t)$ is called the hazard ratio (HR). The main assumption underlying the PH model is that every individual has a proportional hazard.

Accelerated Failure Time model

AFT Models are parametric models that can be used as an alternative to the CPHM in case the proportional hazard assumption is not followed. Unlike the PH model, the main assumption underlying the AFT model is that the effect of a factor is to accelerate or decelerate the time to event by some constant (Klein & Moeschberger, 2003). One of the key advantages of the AFT model is its ability to accommodate a variety of underlying probability distributions for the time to event, such as the Weibull, Exponential, log-logistic, and log-normal distributions (Liu et al., 2018). The Weibull distribution, in particular, is a popular choice as it is both accelerated and proportional, allowing for the simultaneous description of treatment effects in terms of both hazard ratios and event time ratios (Carroll, 2003).

The parametric Accelerated Failure Time (AFT) model assumes that an acceleration factor lambda (λ) relates the survival functions derived from two populations, which can be expressed as a function of covariates. (In this study, two populations refer to the population having the characteristics and the population in the absence of characteristics)

$$S(t) = S_0(t/\lambda(x))$$

Where $S_0(t)$ denotes the baseline survival function, i.e., survival function in the absence of any covariates, and $\lambda(x)$ is the acceleration factor for a group with covariates ($x_1, x_2, x_3......x_p$) is given by

$$\lambda(x) = \exp(\alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_p x_p)$$

The AFT model can accelerate or decelerate the time to event based on the nature of the covariates. The coefficients of the AFT model can be interpreted easily: a unit increase in a covariate means the average survival time changes by a factor of exp (-α_i).

The relation between survival time and the covariates can also be described in the linear form as

$$Y = \log(T) = \mu + \beta_1 x_1 + \beta_2 x_2 ...... + \beta_p x_p + \sigma \in$$

Where μ is the intercept, $\sigma > 0$ is a (unknown) scale parameter, $\beta_1, \beta_2, ...\beta_p$ are the regression coefficients ($\beta_i = -\alpha_i$), and $\in$ is an error distribution, which is an RV assumed to follow a certain distribution. The title of the AFT distribution originates from the distribution that *T* follows instead of the distribution that ln(*T*) follows. The frequently used (parametric) distributions corresponding to the AFT models are Exponential, Weibull, log-normal, and log-logistic distributions. However, in this paper, we have considered Exponential, Weibull, and log-logistic AFT survival models.

Weibull AFT model

Suppose that time *T* (Menstruating Interval to First Conception in our study) follows the Weibull distribution with scale parameter η and shape parameter γ under the AFT model

$$S(t) = \exp(-\eta t^{\gamma})$$

Where $\gamma = 1/\sigma$ and $\eta = \exp(-X_i\beta)$ (acceleration factor by which the time is increasing or decreasing in the presence of the covariate) and the hazard function can be written as

$$h(t) = \gamma\eta t^{\gamma-1}$$

Exponential AFT model

If the value of the shape parameter (i.e. $\gamma = 1/\sigma$) is equal to one, then in that case, the Weibull distribution brings down to the Exponential distribution with survival function $S(t)$ and hazard functions $h(t)$ as follows.

$$S(t) = \exp(-\eta t)$$

$$h(t) = \eta = \exp(-X_i\beta)$$

Log-logistic AFT model

If the error term $(\in)$ supersedes the Loglogistic distribution, then survival and hazard functions of log-logistic AFT distribution are given as;

$$S(t) = [1 + (\eta t)^{1/\gamma}]^{-1}$$

$$h(t) = \frac{\eta^{1/\gamma} t^{((1/\gamma)-1)}}{\gamma\{1 + (\eta t)^{1/\gamma}\}}$$

In case the shape parameter (λ) of the log-logistic distribution is less than or equal to 1, the hazard rate is of a monotonically decreasing nature. In case $\lambda > 1$, it increases from 0 to a maximum and then decreases to zero. More detailed discussions of AFT models are available in textbooks (Collett, 2023; Cox, 2018; Lawless, 2011)

Estimation procedure of an AFT model

The parameter of an AFT model can be estimated with the maximum likelihood estimate (MLE). The likelihood of n observed survival times $t_1, t_2, t_3, \ldots, t_n$ for the log-linear form of AFT distribution is defined as follows

$$L(t, \mu, \beta, \sigma) = \prod_{i=1}^{n} \left[f_i(t_i) \right]^{\delta_i} \left[S_i(t_i) \right]^{1-\delta_i}$$

Where $f_i(t_i)$ and $S_i(t_i)$ are the distribution function and survival function at the time t_i , respectively and δ_i is an indicator variable that takes value 1 if t_i is observed and equal to 0 if

t_i is censored. By taking the logarithm of the above equation and some approximation, the log-likelihood function is

$$\log(L)=\sum_{i=1}^{n}\{-\delta_i\log(\sigma)+\delta_i\log f_{\in_i}(z_i)+(1-\delta_i)\log S_{\in_i}(z_i)\}$$

Where, $z_i=\dfrac{\log t_i-\mu-\beta_1x_1-\beta_2x_2-.....-\beta_px_p}{\sigma}$ and $\in_i$ is a random variable (error term). The estimate of unknown parameters that maximize the likelihood can be estimated by using the Newton-Rhapson method.

The graphical method to check the PH assumption

The Proportional hazards (PH) assumption signifies that the hazard function $h(t)$ of one unit is proportional to that of the hazard function of another unit, i.e., the hazard ratio (HR) is constant over time. The Cox PH survival function can be expressed as

$$S(t,x)=S_0(t)^{\exp(\Sigma\beta_ix_i)}$$

Where $x_i=(x_1,x_2,....x_n)$ are the values of the covariates for an individual. Taking logarithms on both sides twice, we will get

$$ln[-\ln S(t,x)]=\ln[-\ln S_o(t)]+\sum\beta_ix_i$$

Now, for two different individuals having the set of covariates as x_1 & x_2 the difference in the log-log curve is given as

$$\ln[-\ln S(t,x_1)]-\ln[-\ln S(t,x_2)]=\Sigma\beta_i(x_{1i}-x_{2i})$$

Which is independent of *t*. With the help of this relationship, we can show that by plotting the estimated $\ln[-\ln(Survival)]$ versus analysis time for two different groups, we will get parallel curves if the hazard functions are proportional (Hess, 1995). This method is unsuitable for continuous predictors and categorical predictors with numerous categories, as it results in a "cluttered" graph.

The goodness of fit test for checking the PH assumption

The GOF Test based on Schoenfeld residuals is used to evaluate the proportional hazards assumption, which states that the hazard ratios are constant over time. For a given covariate *Xi*, the Schoenfeld residual at time *t* is the difference between the observed covariate value and the expected covariate value under the model. This test involves checking if the residuals show any systematic pattern over time. If the residuals are randomly scattered around zero with no apparent trend, the PH assumption is likely valid. This test examines the null hypothesis

H_0 = The covariate and the time are not correlated.

If the null hypothesis H_0 is rejected [p value ≤ significance level (e.g., .05)], it means that there is evidence against the proportional hazards assumption and the effect of the covariate on the hazard rate changes over time.

Concordance index

The concordance index is often used in survival analysis to evaluate the accuracy of predictive models. It measures how well a model predicts the order of events rather than the exact values of these events. Suppose you have a set of N individuals for each pair of individuals i and j (where i and j are not censored simultaneously). In that case, you check if the individual with the higher predicted risk experiences the event earlier than the other. The C-index is computed as:

$$C = \frac{Number\ of\ concordant\ pairs}{Number\ of\ comparable\ pairs}$$

The C-index ranges between 0.5 and 1. $C = 1$ implies Perfect prediction; the model correctly ranks all pairs. $C = 0.5$ implies the model performs no better than random chance. $C < 0.5$ indicates inverse prediction, where the model is consistently incorrect in the ranking (Harrell et al., 1996).

Comparison of the models

Akaike Information Criteria (AIC) (Akaike, 1974) is used to choose the best-fitted model. The AIC value is determined as the deviance (i.e., -2 times the log-likelihood) plus two times the number of model parameters. The AIC of the model can be calculated as

$$AIC = -2 \times \log(L) + 2k$$

Where k represents the number of estimated parameters in the model, and L denotes the maximum likelihood. The model having the lowest AIC value gives the best-fitted model (Akaike, 1974).

Bayesian Information Criteria serves as a generic model selection tool used to assess the goodness of fit of a model and to compare the relative performance of different models. The BIC is calculated using the following formula:

$$BIC = -2 \times \ln(L) + k \times \ln(n)$$

Where L is the maximum likelihood of the model, k is the number of parameters in the model. n is the number of data points. The model with the lowest BIC is generally preferred (Schwarz, 1978).

Result

The study used information from the 57,446 ever-married women aged 15–44 from NFHS–5 (2019–2021) data. Of these, 9.12% (5,239) are censoring for the menstruating interval to first conception. The censoring information must not be more than 50% in order to achieve acceptable discrimination amongst the parametric models.

Table 1: Percentage Distribution of Socio-Demographic Characteristics of the Respondents

Variable	Category	Frequency (%)	*Tri-mean*	*SIQR*
Respondent's age	< 20	1,313 (2.29)	21.0	13.0
	20–24	9,201 (16.02)	15.8	10.5
	25–29	12,425 (21.63)	13.0	9.0
	≥ 30	34,507 (60.06)	10.8	6.5
Age at marriage	10–15	10,706 (18.60)	29.5	17.0
	16–20	33,713 (58.70)	17.3	11.5
	≥ 20	13,027 (22.68)	13.8	9.5
Place of residence	Urban	10,618 (18.50)	18.6	12.0
	Rural	46,828 (81.50)	15.8	10.5
Religion	Hindu	48,736 (84.80)	18.5	12.0
	Muslim	8,710 (15.16)	15.8	10.5
Caste	SC & ST	15,001 (26.11)	18.5	12.0
	OBC	31,384 (54.60)	18.5	12.0
	UR	11,061 (19.30)	16.5	11.0
Respondent's education	No education	21,445 (37.33)	21.0	13.0
	Primary	7,453 (12.90)	19.8	12.5
	Secondary	20,794 (36.20)	16.5	11.0
	Higher	7,754 (13.50)	14.3	9.5
Mass Media Exposure	Not at all	21,270 (37)	20.5	13.0
	Exposed	36,176 (63)	17.3	11.5
Wealth Index	Poorest	13,534 (23.60)	21.0	13.0
	Poorer	14,545 (25.30)	19.3	12.5
	Middle	10,991 (19.10)	18.3	11.5
	Richer	9,200 (16.00)	16.3	11.5
	Richest	9,176 (15.90)	14.8	9.5
Total	Total females	57,446 (100)	18.5	12

Table 1 presents the distribution of the respondents in the sample for different socio-demographic characteristics and their summary measures. From the table, it is observed that almost 60% (34,507) of women in the study were from the age group 30–44, and around 21.63% (12,425) women belonged to the age group 25–29 at the state level. The tri-mean for MIFC is found to be 21, 16, 13, and 11 months for females in age groups less than 20, 20–24, 25–29, and more than 30, respectively, indicating a negative trend in age and MIFC. 18.6% (10,706) out of the total sample of females in the age group 10–15 are married and thus included in the study as per the given dataset. The majority of the women (58.7%) are married between the ages of 16–20, and approximately 23% of females are married between the age of 20 to 25.

The value of the tri-mean for MIFC is found to be very high at 29.5 months for the females married below the age of 15 which is an obvious result as they have to wait until attaining full biological maturity. The tri-mean of MIFC for age at marriage 16–20 and 20 above is 17 and 14 months, respectively, showing a declining trend between the two. The study revealed that a greater number (82%) of participants are from rural areas, and the rest 18% are from urban areas. The tri-mean for MIFC for urban females is 18.6 months, whereas for rural females, it is 15.8 months, implying that rural females are more likely to conceive earlier than urban females. More than half (55%) of the women are from the OBC Category, and the females in the SC & ST (combined) and UR are approximately 26% and 19%.

Most of the respondents (85%) are from the Hindu religion, and the remaining 15% are followers of the Muslim religion; the tri-mean for MIFC for Hindu and Muslim religions are 18.5 and 15.8, respectively. In our sample, around 63% of women have access to at least one form of the mass media, and only 37% of women are not exposed to any form. The tri-mean for MIFC is observed as 20.5 and 17.3 months for the media-exposed and non-exposed groups. From the table, it is observed that most of the women are from poor backgrounds (24% from the poorest and 25% from poorer), and the wealth index of the females has a negative relation with MIFC. Around 37.33% of the women in the study had no formal education, only 13% (7,453) women had primary education, 36% (20,794) women had secondary education at the state level, and a few percent (13.5%) of females had higher education.

The tri-mean for MIFC is observed as 21, 19.8, 16.5, and 14.3 months for women having no education and having education up to primary, secondary, and higher levels. From the result, it is evident that as the education level of respondents increases the duration between marriage to first conception decreases. Also, the values of spread indicate that the variation in MIFC is high among the women having no education and primary education and decreases further as the education increases. The tri-mean for all the females included in the study is, on average, 18.5 months, i.e., the waiting time for first conception, considering all the females in the study is approximately 1.5 years, and the value of SIQR is 12.

Table 2: Result of Cox's Proportional Hazard Analysis

Variable	Category	Hazard Ratio (HR)	*p* value
Respondent's age	< 20	RC	
	20–24	1.949	.000
	25–29	2.168	.000
	≥ 30	1.889	.000
Age at marriage	< 15	RC	
	15–20	1.449	.000
	≥ 20	1.591	.000
Place of residence	Urban	RC	
	Rural	0.958	.001
Religion	Hindu	RC	
	Muslim	1.103	.000

Variable	Category	Hazard Ratio (HR)	p value
Caste	SC & ST	RC	
	OBC	0.975	.023
	UR	0.987	.037
Respondent's education	No education	RC	
	Primary	0.995	.073
	Secondary	0.926	.019
	Higher	0.988	.048
Mass Media Exposure	Not at all	RC	
	Exposed	1.026	.010
Wealth Index	Poorest	RC	
	Poorer	1.052	.000
	Middle	1.095	.000
	Richer	1.126	.000
	Richest	1.211	.000

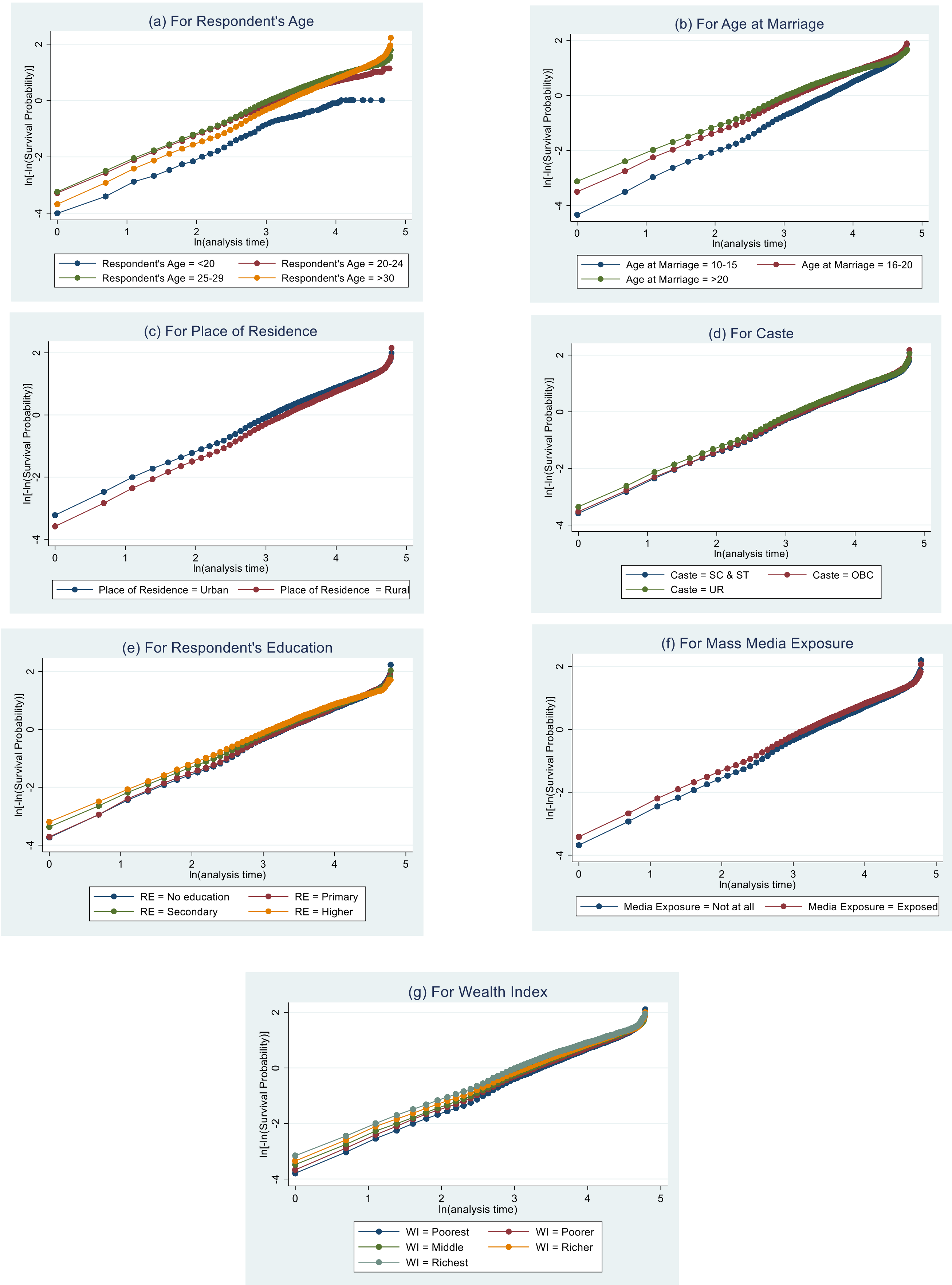
Table 2 shows the result of Cox's proportional hazard model, and Table 3 shows whether the proportionality test is held or violated for the covariates. The result in Table 3 based on Schoenfeld residual GOF test shows that the assumption of proportional hazard is violated for all the covariates as the p value is less than .05 (significance level).

Table 3: The Goodness of Fit (GOF) Test Based on Schoenfeld Residual for the Variables

Variable	HR	SE	p value
Respondents age	1.135	0.013	.000
Age at marriage	1.223	0.007	.000
Place of residence	0.844	0.009	.000
Religion	1.136	0.013	.000
Caste	1.030	0.004	.015
Respondent's education	1.063	0.004	.000
Mass Media Exposure	1.127	0.010	.000
Wealth Index	1.075	0.003	.000

The PH assumption is graphically accessed and presented in Figure 1. The graphical procedure involves plotting *log[-log(survival)]* versus *log (analysis time*) for all the groups in every categorical variable where survival function $S(t)$ and survival time t are at a given time t for different covariates. The plots do not yield parallel curves and clearly show the proportional hazard assumption is not followed by any of the covariates. These graphs are constructed by using the 'survival' package in Stata version 16. The value of the concordance index is 0.57 which also shows that the fitting of the Cox PH model is poor. So, this article further used the parametric AFT model to determine the significant factors affecting the waiting time for the first conception in Uttar Pradesh.

Figure 1: The Log Survival Plots to Assess the PH Assumption of Categorical Covariates



Result of parametric AFT models

Three parametric accelerated failure time models, Exponential AFT distribution, Weibull AFT distribution, and Loglogistic AFT distribution, are used in this study to identify the significant covariates. Table 4 presents the results of three AFT models. The time ratio (TR) of the model provides a multiplicative factor by which the survival time is accelerated or decelerated due to the presence of covariates. The STATA version 16 software is used to perform the analysis. The parameter estimate and its corresponding *p* value of AFT Models are given in Table 4.

Table 4: Comparison between the Cox PH Model and Different AFT Distributions using Akaike Information Criteria

Models	Log-likelihood	AIC	BIC
Cox PH	-520,749.4	1,010,083	1,010,244
Exponential AFT	-82,424.42	155,441.4	155,620.1
Weibull AFT	-81,919.39	153,172.7	153,360.3
Loglogistic AFT	-83,943.23	155,966.8	156,154.3

The selection of models in this study was based on the assumed distribution of the baseline hazard. Given this assumption, comparisons will be made among parametric AFT models if the baseline hazard follows a specific distributional pattern. In this study, the baseline hazard exhibited a distributional pattern. It is observed from Table 5 that among the three AFT models (Exponential AFT, Weibull AFT, and Loglogistic AFT), the AIC estimate of the Weibull AFT model is the lowest. Hence, we have contemplated Weibull AFT as our working model.

The estimate of the Weibull AFT model shows that the current age of females, age at first marriage, place of residence, religion, caste, educational level of respondents, wealth status, and mass media exposure of women have a statistically significant relationship with time to first conception. From Table 4, The waiting time for the first conception decreases as the woman's age at marriage increases, i.e., the first conception occurs more quickly for those women whose age at marriage is high.

According to the result, younger women have a large waiting time up to first conception. However, after the age of 30 years, the transition time towards the first union has increased compared to the age group of 20 to 30 years. Women from rural backgrounds (TR=0.944) have a faster transition towards first union as compared to women from urban backgrounds.

The women having primary (TR = 1.007), secondary (TR = 1.079), and higher (TR = 1.305) education take more time before their first conception than the women having no education. Women coming from the Hindu religion have a longer duration of waiting time to first conception compared to Muslim women (TR = 0.902). The increased waiting time to first conception was found for women having UR (TR = 1.404) and OBC (TR = 1.029) categories than women having SC & ST categories.

Exposure to mass media (TR = 1.073) significantly increases the waiting time for the first conception, thus delaying family planning. Again, the above result revealed that women's wealth status significantly increased their waiting time until their first union.

Table 5: Result of Parametric Accelerated Failure Time Models

Variable	Category	Exponential AFT			Weibull AFT			Loglogistic AFT		
		Time Ratio	*p* value	St. Error	Time Ratio	*p* value	St. Error	Time Ratio	*p* value	St. Error
Respondents age	< 20					RC				
	20–24	0.578	.000	0.026	0.530	.000	0.021	0.536	.000	0.024
	25–29	0.528	.000	0.021	0.482	.000	0.020	0.496	.000	0.022
	≥ 30	0.640	.000	0.020	0.585	.000	0.018	0.636	.000	0.020
Age at marriage	≤ 15					RC				
	16–20	0.703	.000	0.008	0.688	.000	0.007	0.643	.000	0.007
	≥ 20	0.645	.000	0.009	0.623	.000	0.008	0.556	.000	0.008
Place of residence	Urban					RC				
	Rural	0.944	.001	0.013	0.946	.000	0.012	0.963	.000	0.014
Religion	Hindu					RC				
	Muslim	0.912	.000	0.012	0.902	.000	0.011	0.867	.000	0.012
Caste	SC &ST					RC				
	OBC	1.024	.027	0.011	1.029	.004	0.010	1.047	.000	0.012
	UR	1.309	.052	0.014	1.404	.027	0.013	1.219	.012	0.015
Respondent's education	No education					RC				
	Primary	1.007	.061	0.014	1.002	.082	0.013	1.002	.084	0.014
	Secondary	1.079	.026	0.010	1.064	.001	0.010	1.063	.000	0.010
	Higher	1.305	.035	0.016	1.302	.099	0.015	1.415	.042	0.016
Mass Media Exposure	Not at all					RC				
	Exposed	1.073	.010	0.010	1.072	.004	0.009	1.063	.001	0.010
Wealth Index	Poorest					RC				
	Poorer	1.050	.000	0.012	1.051	.000	0.011	1.058	.000	0.012
	Middle	1.092	.000	0.013	1.092	.000	0.012	1.096	.000	0.013
	Richer	1.121	.000	0.014	1.122	.000	0.013	1.136	.000	0.015
	Richest	1.200	.000	0.015	1.201	.000	0.014	1.213	.000	0.015

Discussion

Investigating the duration between marriage and first conception is crucial, as it signifies a woman's initiation into motherhood and reflects the family planning situation. The study found that as the age of females increases, the menstruating interval to first conception increases up to the age of 30. However, females aged more than 30 years have larger menstruating intervals than women aged 20 to 30 years. This might be due to factors such as occupation or infertility, which might need further investigation as we do not have such information in our data.

Age at marriage also seemed to be a potential determinant in this study. Females who marry late have a small waiting time, and it is obvious that mature women do not have to wait much before entering into motherhood, which is also supported by many studies (Nath, 1993; Obite et al., 2021). The respondents' place of residence has a significant impact on the MIFC in which the rural females have less time before their first union. The findings revealed that the women who were exposed to mass media had experienced prolonged waiting times, which is also sustained by the study of (Rahman et al., 2013). This result is quite satisfactory, as women with media exposure tend to gain more knowledge about maintaining a healthy family life compared to those without media access.

According to the study, the respondent's education had a significant impact on the MIFC. We have found that respondents having primary, secondary, and higher education have a longer waiting time than the women having no education. The main reason behind this scenario might be the more exposure and more awareness towards family planning of the higher educated women. This finding is similar to the study of Drèze and Murthi (2001).

Religion is a covariate affecting the survival of time to the first conception of women. Muslim and other religious follower women have shorter survival time to first conception than women from the Hindu religion. This outcome is still consistent with studies conducted in rural northern India by Yadava et al. (2000).

We also observed that from the Weibull AFT model, respondents from higher wealth status have significantly longer waiting times for first conception than respondents having poor backgrounds; a similar result is obtained by Singh et al. (2016).

In this study, we have identified some key indicators that are affecting the Waiting time for first conception. However, these are only some of the indicators that decide the duration of waiting time to the first conception due to the limitation of data availability. Analyzing MIFC helps identify population-wide reproductive health trends, which may vary based on factors like age, socioeconomic conditions, or exposure to environmental stressors. Understanding these factors affecting MIFC will help policymakers and healthcare providers design interventions to improve fertility outcomes. For instance, if a specific part of the population exhibits prolonged MIFC due to delayed marriage or access to reproductive healthcare, policies addressing these issues can be implemented.

Public health strategies promoting optimal MIFC through awareness programs, the use of contraception, and healthcare access are essential for improving global health metrics. Targeted interventions are also essential in these areas to address particular sections of society. Policy efforts emphasizing family planning, female education, and accessibility will significantly contribute to reducing the risks associated with short MIFC.

Conclusion

Although the Cox proportional model with its related form and extensions is still crucial for analyzing censored survival data, it is still necessary to show that other approaches are available that can be used for other additional and important information of data that the Cox PH model is not intended to analyze. In this article, the semiparametric CPH model and parametric AFT distributions were practiced on the birth interval data in Uttar Pradesh, India, which still has the TFR above the replacement level, making it important to revisit the factors affecting the MIFC. In many real-life problems, researchers prefer to work with the CPHM rather than the AFT models without even investigating the necessary PH assumption (Altman et al., 1995).

The violation of the PH assumption can lead to the Cox proportional hazard model being unreliable and biased (Zare et al., 2015). Based on the goodness of fit (GOF) test (based on Schoenfeld residual), it is visible that the proportional hazard assumption for the birth interval data in Uttar Pradesh is violated. In the case of non-PH behavior in the data set, we have applied three parametric AFT models to the BI data. Furthermore, numerous studies have exemplified that AFT models are worthier than PH models for fitting the data with non-PH assumptions. (Saikia et al., 2017; Wei, 1992).

The AIC value of models indicates that the Weibull AFT distribution had the smallest AIC value compared to the CPHM, Exponential AFT distribution, and Loglogistic AFT distribution, implying that the Weibull AFT is the well-fitted distribution for the birth interval data in Uttar Pradesh among all the considered models. Moreover, the AFT model also offers a distinct advantage when the research focuses on directly estimating hazard functions rather than solely on the relative risk (Moolgavkar et al., 2017).

The CPH model, while versatile, may struggle to accurately capture the effects of time-varying covariates, leading to potential biases in the estimation of hazard ratios. In contrast, the AFT model allows for more flexible modeling of the temporal aspects of exposure and risk, enabling researchers to gain deeper insights into the association between covariates and the time-to-event outcome (Cox et al., 2007).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/tac.1974.1100705

Altman, D. G., De Stavola, B. L., Love, S. B., & Stepniewska, K. A. (1995). Review of survival analyses published in cancer journals. *British Journal of Cancer*, *72*(2), 511–518. https://doi.org/10.1038/bjc.1995.364

Bakhshi, E., Ali Akbari Khoei, R., Azarkeivan, A., Kooshesh, M., & Biglarian, A. (2017). Survival analysis of thalassemia major patients using Cox, Gompertz proportional hazard and Weibull accelerated failure time models. *Medical Journal of the Islamic Republic of Iran, 31*(1), 568–571. http://dx.doi.org/10.14196/mjiri.31.97

Carroll, K. J. (2003). On the use and utility of the Weibull model in the analysis of survival data. *Controlled Clinical Trials*, *24*(6), 682–701. https://doi.org/10.1016/S0197-2456(03)00072-2

Chandra-Mouli, V., Camacho, A. V., & Michaud, P. A. (2013). WHO guidelines on preventing early pregnancy and poor reproductive outcomes among adolescents in developing countries. *Journal of Adolescent Health*, *52*(5), 517–522. https://doi.org/10.1016/j.jadohealth.2013.03.002

Collett, D. (2023). *Modelling survival data in medical research*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003282525

Cox, C., Chu, H., Schneider, M. F., & Munoz, A. (2007). Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine*, *26*(23), 4352–4374. https://doi.org/10.1002/sim.2836

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Cox, D. R. (2018). *Analysis of survival data*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315137438

Drèze, J., & Murthi, M. (2001). Fertility, education, and development: Evidence from India. *Population and Development Review*, *27*(1), 33–63. https://doi.org/10.1111/j.1728-4457.2001.00033.x

Faruk, A. (2018). The comparison of proportional hazards and accelerated failure time models in analyzing the first birth interval survival data. *Journal of Physics: Conference Series, 974*, Article 012008. https://doi.org/10.1088/1742-6596/974/1/012008

Ghimire, D. J., & Axinn, W. G. (2010). Community context, land use, and first birth. *Rural Sociology*, *75*(3), 478–513. https://doi.org/10.1111/j.1549-0831.2010.00019.x

Harrell, F. E., Jr., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine, 15*(4), 361–387. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

Hess, K. R. (1995). Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. *Statistics in Medicine, 14*(15), 1707–1723. https://doi.org/10.1002/sim.4780141510

International Institute for Population Sciences (IIPS) & ICF. (2022, March). *National Family Health Survey (NFHS-5), 2019–21: India: Volume I*. https://dhsprogram.com/pubs/pdf/FR375/FR375.pdf

Kalbfleisch, J. D., & Prentice, R. L. (2011). *The statistical analysis of failure time data*. John Wiley & Sons.

Kay, R., & Kinnersley, N. (2002). On the use of the accelerated failure time model as an alternative to the proportional hazards model in the treatment of time to event data: a case study in influenza. *Drug Information Journal*, *36*(3), 571–579. https://doi.org/10.1177/009286150203600312

Khanal, S. P., Sreenivas, V., & Acharya, S. K. (2014). Accelerated failure time models: An application in the survival of acute liver failure patients in India. *International Journal of Science and Research, 3*(6), 161–166. https://doi.org/10.21275/201445

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (Vol. 1230). Springer. http://dx.doi.org/10.1007/b97377

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*. John Wiley & Sons. https://doi.org/10.1002/9781118033005

Liu, E., Liu, R. Y., & Lim, K. (2023). Using the Weibull accelerated failure time regression model to predict time to health events. *Applied Sciences*, *13*(24), Article 13041. https://doi.org/10.3390/app132413041

Mendes, A. C., & Fard, N. (2014). Accelerated failure time models comparison to the proportional hazard model for time-dependent covariates with recurring events. *International Journal of Reliability, Quality and Safety Engineering*, *21*(02), Article 1450010. https://doi.org/10.1142/S0218539314500107

Moolgavkar, S. H., Chang, E. T., Watson, H. N., & Lau, E. C. (2018). An assessment of the Cox proportional hazards regression model for epidemiologic studies. *Risk Analysis*, *38*(4), 777–794. https://doi.org/10.1111/risa.12865

Nath, D. C., Land, K. C., & Singh, K. K. (1995). A waiting time distribution for the first conception and its application to a non-contracepting traditional society. *Genus*, 95–103. https://www.jstor.org/stable/29789186

Nath, D. C., Singh, K. K., Land, K. C., & Talukdar, P. K. (1993). Age of marriage and length of the first birth interval in a traditional Indian society: Life table and hazards model analysis. *Human Biology*, 783–797. https://www.jstor.org/stable/41464906

Obite, C. P., Bartholomew, D. C., Nwosu, U. I., Anyiam, K. E., & Aminu, S. A. (2021). Marriage to first birth interval in Nigeria: Analysis of the roles of social-demographic and cultural factors. *SN Social Sciences*, *1*(5), Article 120. https://doi.org/10.1007/s43545-021-00112-x

Orbe, J., Ferreira, E., & Núñez-Antón, V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in Medicine*, *21*(22), 3493–3510. https://doi.org/10.1002/sim.1251

Páez, A., & Boisjoly, G. (2022). Exploratory data analysis. In *Discrete choice analysis with R* (pp. 27–72). Springer. https://doi.org/10.1007/978-3-031-20719-8_2

Qi, J. (2009). *Comparison of proportional hazards and accelerated failure time models* (Master's thesis, University of Saskatchewan). HARVEST Institutional Repository. http://hdl.handle.net/10388/etd-03302009-140638

Rahman, M., Mustafi, M., & Azad, M. (2013). Analysis of the determinants of marriage to first birth interval in Bangladesh. *International Journal of Management and Sustainability*, *2*(12), 208–219. https://doi.org/10.18488/journal.11/2013.2.12/11.12.208.219

Saikia, R., & Barman, M. P. (2017). A review on accelerated failure time models. *International Journal of Statistics and Systems, 12*(2), 311–322. https://www.ripublication.com/ijss17/ijssv12n2_14.pdf

Schemper, M. (1992). Cox analysis of survival data with non-proportional hazard functions. *Journal of the Royal Statistical Society Series D: The Statistician*, *41*(4), 455–465. https://doi.org/10.2307/2349009

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, *69*(1), 239–241. https://doi.org/10.1093/biomet/69.1.239

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Sheps, M. C., Menken, J. A., Ridley, J. C., & Lingner, J. W. (1970). Truncation effect in closed and open birth interval data. *Journal of the American Statistical Association*, *65*(330), 678–693. https://doi.org/10.1080/01621459.1970.10481116

Singh, B. P., Singh, G., & Singh, K. K. (2016). Heterogeneity in waiting time to first conception in Uttar Pradesh: A parametric regression analysis. *Demography India, 45*(1/2), 77–88. https://iasp.ac.in/uploads/journal/006-1635845884.pdf

Swindell, W. R. (2009). Accelerated failure time models provide a useful statistical framework for aging research. *Experimental Gerontology*, *44*(3), 190–200. https://doi.org/10.1016/j.exger.2008.10.005

Talwar, P. P. (1967). A note on changes in age at marriage of females and their effect on the birth rate in India. *Eugenics Quarterly*, *14*(4), 291–295. https://doi.org/10.1080/19485565.1967.9987739

Tripathy, M. R., Swain, P. K., Sarangi, P. K., & Pattnaik, S. S. (2022). Accelerated Failure Time Models with Applications to Endometrial Cancer Survival Data. *Journal of Reliability and Statistical Studies*, 723–744. https://doi.org/10.13052/jrss0974-8024.15213

Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine, 11*(14–15), 1871–1879. https://doi.org/10.1002/sim.4780111409

Wilcox, R. R. (2012). *Introduction to robust estimation and hypothesis testing*. Academic press.

Wood, J. W. (2017). *Dynamics of human reproduction: Biology, biometry, demography*. Routledge. https://doi.org/10.4324/9780203792780

Yadava, K. N. S., Jain, S. K., & Kumar, A. (2000). Correlates of birth intervals in rural Northern India. *Genus, 56*(1/2), 55–76. https://www.jstor.org/stable/29788629

Zare, A., Hosseini, M., Mahmoodi, M., Mohammad, K., Zeraati, H., & Holakouie-Naieni, K. (2015). A comparison between accelerated failure-time and Cox proportional hazard models in analyzing the survival of gastric cancer patients. *Iranian Journal of Public Health, 44*(8), 1095–1102. https://ijph.tums.ac.ir/index.php/ijph/article/view/3359

Zhang, Z. (2016). Parametric regression model for survival data: Weibull regression model as an example. *Annals of Translational Medicine,* 4(24), Article 484. https://doi.org/10.21037/atm.2016.08.45