Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2024;10(2): 192-208

Original Article

Adjustment of Truncation Effect in First Birth Interval using Current Status Data Technique

Sachin Kumar[1], Anup Kumar[1*], Amit Kumar Misra[2], Jai Kishun[1], Uttam Singh[1]

[1]Department of Biostatistics and Health Informatics Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India.
[2]Department of Statistics, Babasaheb Bhimrao Ambedkar University Lucknow, India.

ARTICLE INFO

ABSTRACT

Received 26.03.2024
Revised 09.04.2024
Accepted 11.04.2024
Published 15.06.2024

Key words:
Spline smoothing;
Truncation effect;
Selection bias;
Current status data;
Ayer method

Introduction: Estimating the First Birth Interval (FBI) from cross-sectional data often presents challenges related to truncation effects. These challenges stem from the data's inability to capture the enough exposure for an event, resulting in potential biases and inaccuracies in FBI estimates. Recognizing and addressing truncation effects is essential for obtaining more precise and meaningful fertility parameter estimates in a cross-sectional survey.
This study seeks to mitigate truncation effects in the estimation of the FBI by utilizing the Current Status Data technique. This approach focuses on women with specific marital durations, providing a means to counteract the bias caused by truncation and thereby yielding more accurate and reliable FBI estimates.
Methods: Data from the National Family Health Survey (NFHS-IV) are employed for this study. The Current Status Data Technique is applied to the dataset, considering exclusively those women with marital durations less than 120 months. This methodology enables the adjustment of truncation effects and facilitates a more precise estimation of the FBI. Statistical analysis is conducted to determine the FBI distribution and ascertain the necessary sample size.
Results: The estimated First Birth Interval (FBI) without accounting for truncation is 27.85 months, while the estimate considering truncation is 31.70 months. When applying the Current Status Data technique, the estimated FBI is 30.70 months. To obtain reliable estimates of the FBI using Current Status techniques, a minimum sample size of over 5,000 observations is necessary.
Conclusion: The truncation effect in FBI is addressed, and some non-parametric adjustments are used for estimating the duration of FBI. The Current Status Data technique emerges as a valuable tool for mitigating these effects and enhancing the precision of FBI estimates. This research contributes to an improved understanding of fertility dynamics and provides valuable insights for future studies on the First Birth Interval.

Introduction

Estimating the distribution of the First Birth Interval (FBI) ideally involves tracking a cohort of women until all have completed their first birth. However, such data is often unavailable and difficult to collect. Instead, cross-sectional data from national health surveys, though

*.Corresponding Author: anup.stats@gmail.com

practical, require statistical adjustments for accurate analysis. These surveys are widely used to study event durations like breastfeeding, postpartum amenorrhoea (PPA), and birth intervals, with applications in demography and medicine. Key challenges include censoring, selection bias, and truncation, which can introduce bias in estimating event distributions if not properly addressed.

Censoring occurs when individuals do not experience the event of interest within the study period. For example, some women may not have given birth by the time of the survey, leading to censored FBI data. Estimating the average FBI using only complete cases underestimates the true duration, as shown by previous studies.[1, 2] Censoring complicates analysis, requiring survival methods to handle time-to-event data. Techniques like imputation and likelihood-based approaches have been developed over time to analyze censored survival data.[3]

The timing of an event is often linked to the length of exposure. In the case of the FBI, exposure time varies among women, influencing the study's results. Women with longer marital durations at the time of the survey are more likely to have their first birth recorded, while shorter durations often lead to truncated data. Researchers have addressed this truncation issue by focusing on women with longer marital durations (e.g., over 7 or 10 years).[4–12] However, this approach overlooks recently married women, leading to outdated estimates. Studies[12, 13] have explored FBI distribution for shorter marital durations in more homogeneous groups.

Several authors have used parametric approaches to adjust for truncation and selection bias in FBI studies.[1, 14–16] When birth intervals are modeled based on specific marital durations, these adjustments provide more accurate parameter estimates. Models accounting for truncation have been developed accordingly.[17, 18] Additionally,[19] explored the asymptotic properties of Non-Parametric Maximum Likelihood Estimation (NPMLE) in interval-truncated data. Current status data, which arise in various fields like demography, epidemiology, and medicine, are often used in cross-sectional studies to examine events such as first pregnancy.[20–22] In medical studies, FBI and HIV data frequently involve current status data due to the challenge of measuring exact event times.

Current status data is a non-parametric approach used to adjust for truncation and selection bias. It involves cases where the event time T is not observed directly; instead, it is only known whether T occurs before or after a random examination time X.[23, 24] In FBI studies,[17, 25] researchers cannot observe the exact marriage-to-birth duration but only whether it occurred before or after the survey. These observations, categorized as either $(0, X]$ or (X, ∞), are known as Case-I interval-censored data.[26] Since current status data is not based on recall, it can provide unbiased estimates free from truncation and selection bias.

After estimating the distribution using current status data, determining the optimal sample size becomes essential in study design. While sample size determination is well-explored in survival analysis, discussions specific to current status data are limited.[27–29] For instance,[30] proposed a method to design cross-sectional surveys estimating disease incidence based on current age and health status. Traditional sample size methods for right-censored data focus on accuracy and confidence intervals.[27, [28, 31, 32] The focus is on estimating population characteristics with a specified margin of error and confidence interval, acknowledging that sample estimates rarely match the true population parameter exactly. The key is to define how much error is acceptable and ensure the confidence interval reflects the true value with a certain probability.

Section 2 outlines the materials and methods used to estimate FBI distribution with the current status data technique. Section 3 applies these methods to FBI and discusses optimal sample size determination using two approaches. Section 4 presents simulation results and discussion.

Material and Method

Current Status Data and Estimation

This section is devoted on estimation of the distribution of FBI using the current status data of women whose marital duration is less than or equal to T months. A brief description of methodology on current status data in cross-sectional study is given as.[33, 34] In a cross-sectional survey, the marital duration and the status of first birth is noted. Let $(T_1, \delta_1),(T_2, \delta_2),\ldots,(T_n, \delta_n)$ be n pairs of random variables, where T_i and δ_i, respectively, denote the marital duration and the status of first birth of i^{th} woman at the time of survey, i=1,2,…,n; and

$$\delta_i = \begin{cases} 0, \text{if} X_i > T_i \\ 1, \text{if} X_i \le T_i \end{cases}$$

where X_i is the non-negative random variable representing the FBI of the i^{th} woman, i=1,2,…,n. Let the cumulative distribution function of X_i be $F_X(.)$,i=1,2,…,n. It is also assumed that X_i is independent of T_i,i=1,2,…,n. Our aim is to estimate $F_X(t)$ for all $t \ge 0$.

Let $0 = t_0 < t_1 < t_2 < \ldots < t_k = T$ be the observed marital durations. Let n_j be the number of women having marital duration in the interval $(t_{j-1}, t_j]$ months at the time of survey, j=1,2,…,k. Further, let y_j be the number of women with martial duration in the interval $(t_{j-1}, t_j]$ months who don't have the first birth (i.e., the FBI is greater than tj months) at the time of survey, j = 1, 2, . . . , k. Mathematically, if we let

$$A_j=\{i:t_{j-1} \le T_i \le t_j\}$$

and

$$B_j=\{i:t_{j-1} \le T_i \le t_j, \delta_i=0\}, j=1,2,\ldots,k,$$

then

n_j="number of elements in " A_j and y_j="number of elements in " B_j, j=1,2,…,k,

Now, for j=1,2,…,k, we have

$$\begin{aligned} F_X(t_j) &= P(X_1 \le t_j) \\ &= 1 - P(X_1 > t_j) \\ &= 1 - S_X(t_j), \end{aligned}$$

where $S_X(t_j)$ is estimated by

$S(tj) = \frac{y_j}{n_j}$ ="proportion of women whose FBI is greater than " t_j, j=1,2,…,k. (1)

Thus, we have estimates of $S_X(t_1), S_X(t_2),\ldots,S_X(t_k)$ as $\overline{S}(t_1),\overline{S}(t_2),\ldots,\overline{S}(t_k)$. Using these t_k values, we have to estimate $S_X(t)$ for all t≥0 which in turn gives the estimate of F(t),t≥0. In order to get the estimate of $S_X(t)$ for all t≥0, we apply spline smoothing technique.[35] A spline $S_X(.)$, is a smooth, piecewise-defined function composed of low-degree polynomials, each defined on specific intervals of t_j. These polynomial pieces are joined together at points called knots, ensuring the function remains smooth across the entire range of t_j.

By Ayer Method

Using the notation defined in Section 2, we have $\overline{S}(t_j) = \frac{y_j}{n_j}, j = 1, 2, \ldots, k.$

Denote $\overline{S}(t_j) = s_j, j = 1, 2, \ldots, k$.[36] suggested that the maximum likelihood estimates $\hat{s}_j$ of $S(t_j)$ may be found in the following way.

Case I: If $s_1 \geq s_2 \geq \ldots \geq s_k \geq 0$, then $\hat{s}_j = s_j$, j= 1,2,…, k.

Case II: If $s_j < s_{j+1}$ for some j(j=1,2,…,k-1), then take $\hat{s}_j = \hat{s}_{j+1}$. To find $\hat{s}_j$, the ratios

$s_j = y_j / n_j$ and $s_{j+1} = y_{j+1} / n_{j+1}$ are replaced in the sequence $s_1, s_2, \ldots, s_k$ by the single ratio $(y_j + y_{j+1})/(n_j + n_{j+1})$ which gives us an ordered set of only k-1 ratios. This procedure is repeated until an ordered set of ratios is obtained which are monotonic non-increasing. Then, for each j, $\hat{s}_j$ is equal to that one of the final set of ratios to which the original ratio $s_j = y_j/n_j$ contributed. In other words, if for integers r,s, with $1 \leq r \leq s \leq k$, we define

$$\acute{a}(r,s) = \sum_{j=r}^{s} y_j,$$

$$\hat{a}(r,s) = \sum_{j=r}^{s} (n_j - y_j),$$

and

$$A(r,s) = \frac{\acute{a}(r,s)}{\acute{a}(r,s) + \hat{a}(r,s)},$$

and

$$\hat{s}_j = \min_{1 \leq r \leq j} \max_{j \leq s \leq k} A(r,s)$$

Then, the MLE of F(.) is given by

$$\hat{F}(t) = \begin{cases} 0, & \text{if } t < t_1 \\ 1 - \hat{s}_j, & \text{if } t_j \leq t < t_{j+1}, j = 1, 2, \ldots, k-1, \\ 1, & \text{if } t \geq t_k \end{cases}$$

The methodology outlined above serves the purpose of estimating F(t). It entails intricate statistical calculations and assumptions, aiming to derive an estimate of F(t) (as shown in Figure 1). It's worth emphasizing that this estimate is susceptible to the inherent variability and uncertainties commonly encountered in real-world data.

In the quest to assess the reliability and precision of the FBI estimate, the need arises

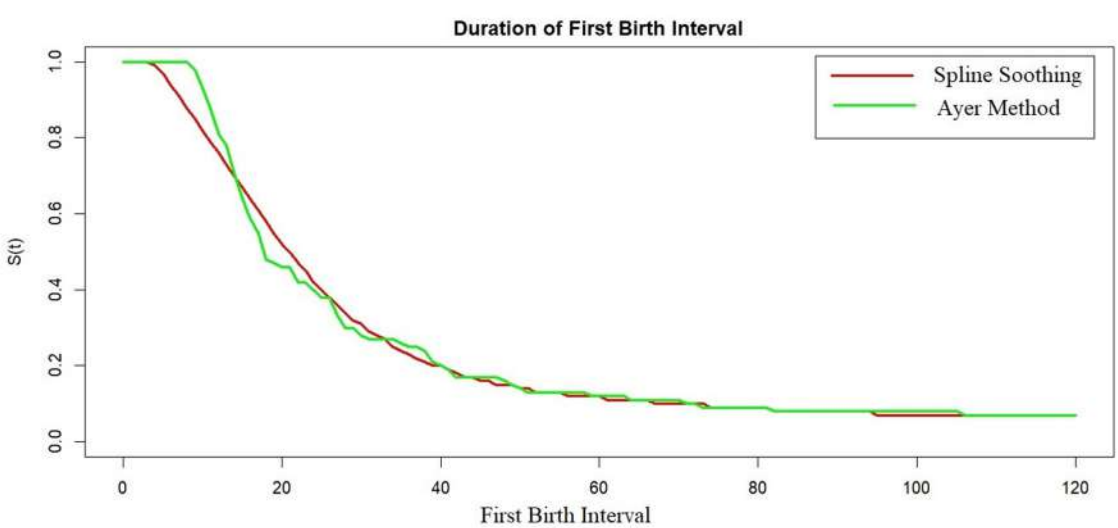

Figure 1. Distribution of FBI with different method

to determine the minimum required sample size. This endeavor involves the calculation of the necessary number of observations or data points essential to achieve a predetermined level of statistical accuracy and confidence in the estimate. This sample size calculation takes into account parameters such as the desired level of statistical significance (referred to as η), the margin of error (ϵ), and various other relevant statistical considerations.

Sample size using Ayer Method

As stated in Theorem 3.1 of [36], let's assume that t_0 represents a point of continuity within the distribution function F(t). Select arbitrary positive values for ϵ (the margin of error) and η (the level of significance) such that when t' and t″ are chosen such that $t' \leq t_0 \leq t''$ and $|F(t)-F(t_0)| \leq \epsilon/2$ for $t' \leq t_0 \leq t''$, then:

$$pr\{|\overline{F}(t_0)-F(t_0)|\leq\epsilon\}>1-\eta$$

This condition holds provided that there are at least N trials conducted between t' and t_0, as well as at least N trials between t' and t″, with N determined such that:

$$\sum_{j=N}^{\infty}\frac{1}{j^2}+\frac{1}{4N}<\frac{\epsilon^2 ç}{32} \qquad (2)$$

This value of N represents the required sample size for your analysis.

Simulation Study for finding Minimum Sample Size

To determine the minimum sample size required for estimating FBI, a simulation approach was utilized. Independent samples of sizes n = 1000, 1500, 2500, 3000, 4000, 5000, and 6000 were drawn from the Individual record (IR) file of NFHS-IV. This sampling procedure was repeated 1000nd times for each sample size. Subsequently, an analysis of the FBI was conducted as discussed in above section. The mean and quartile of the FBI were computed based on 1000nd repetitions, along with a 95% confidence interval, (as presented in Tables-1).

Table 1. Distribution of sample using Spline Smoothing and Ayer method with 1000nd repetition

S. no	S.S	Spline Smoothing				Ayer Method			
		1stQuantile (95% CI)	2nd Quantile / Median (95% CI)	3rdQuantile. (95% CI)	Mean (95% CI)	1stQuantile (95% CI)	2nd Quantile / Median (95% CI)	3rdQuantile. (95% CI)	Mean (95% CI)
1	1000	11.37(10, 13)	20.20(18, 23)	37.37(31, 46)	32.16(27.38,40.27)	12.77(11, 15)	20.80(16, 25)	37.42(29, 51)	29.48(26.96, 1.98)
2	1500	11.42(10, 13)	20.16(18, 22)	36.47(32, 42)	31.40(27.87, 7.89)	12.73(11, 15)	20.59(17, 25)	36.99(29, 48)	29.45(27.40, 1.55)
3	2500	11.41(10, 13)	20.12(19, 22)	36.27(33, 41)	30.62(28.07, 4.94)	12.72(12, 14)	20.58(18, 24)	36.67(30, 46)	29.45(27.98, 1.03)
4	3000	12.09(11, 13)	20.88(20, 22)	37.69(34, 42)	31.42(29.01, 5.57)	13.12(12, 14)	21.14(19, 24)	37.88(31, 46)	30.50(29.12,31.87)
5	4000	12.06(11, 13)	20.80(20, 22)	37.52(34, 1.02)	31.08(29.03,34.68)	13.11(12,14)	21.09(19,24)	37.75(31,46)	30.42(29.15,31.68)
6	5000	11.38(11, 12)	20.08(19, 21)	36.20(33, 39)	29.98(28.31, 3.07)	12.63(12, 13)	20.59(18, 24)	36.36(31, 42)	29.41(28.35,30.48)
7	6000	12.07(11, 13)	20.81(20, 22)	37.55(35, 41)	30.85(29.17, 3.69)	13.08(12, 14)	21.04(19, 23)	37.55(32, 43)	30.42(29.37, 1.45)
8	7500	11.33(11, 12)	20.02(19, 21)	36.19(34, 39)	29.72(28.37, 2.42)	12.66(12, 13)	20.45(19, 23)	36.25(32, 40)	29.39(28.50, 0.31)
Population		12	20	37	30.7	13	21	37	30.36

Result

The First Birth Interval (FBI) estimation without considering truncation effects is 27.85 months. However, when truncation effects are adjusted for, the estimate increases to 31.70 months. By employing the Current Status Data technique, the estimated FBI is refined to 30.70 months. To obtain reliable estimates of the FBI using Current Status techniques, a minimum sample size of over 5,000 observations is necessary.

The study utilizes data of individual record (IR) file of National Family Health Survey (NFHS-IV) [37] of india and considering the variable such as Date of Interview (DoI), Date of Marriage (DoM) and Marriage to Firth interval (in months).

After applying Kaplan-Meier estimation to the cross-sectional data, the mean estimate for the FBI is found to be 27.85 months, (as shown in Table 2). However, these estimates have certain limitation. The most significant limitation is that the length of exposure for each woman who experienced the first birth is varying. Consequently, women with short marital durations are truncated from the study.

To overcome this limitation,[25] proposed an approach to ensure that all women receive appropriate exposure time, with a guaranteed high probability of experiencing their first birth. Only those women whose marital duration exceedes 120 months (i.e., $T_i \geq 120$) are consider. The mean duration of the FBI was computed again resulting in an estimated value of 31.70 months, (as shown in Table-2). But this frame has following drawbacks: (i) The estimates obtain from the women of longer marital duration are not based on recently married women. (ii)The sample size becomes smaller as it excludes considerable number of recently married women. (iii) Due to recall lapse, older women may not report the duration of FBI correctly.

To overcome these limitations and adjustment of truncation effect, the current status data technique is applied only to those women of marital durations within $T_i \leq 120$ months (Table S1 in supplementary). Following the methodology detailed in Section 2, the estimated FBI is 30.36 months with spline smoothing and about 30.70 months using the Ayer Method as shown in Table 2.

The curves generated by Spline Smoothing and the Ayer Method are shown in Figure 1. The Spline Smoothing estimate of the FBI is derived; however, it lacks a specific mathematical form due to its reliance on knots and degrees of freedom. Different combinations of knots and degrees of freedom can produce varying smoothing functions for the FBI. Consequently, determining the pair of knot and degree of freedom that provides the most accurate FBI estimate is challenging.

In contrast, when utilizing the Ayer Method,

Table 2. Descriptive statistics of (First Birth Interval) FBI using two method

Statistics	Kaplan Meier Estimation		Current Status Data	
	Cross-Sectional data	Ti ≥ 10 year	Spline Smoothing	Ayer Method
1st Quartile	13.00	14.00	13.00	12.00
2ndQuartile/Median	21.00	23.00	21.00	20.00
3rd Quartile	33.00	36.00	37.00	37.00
Mean	27.85	31.70	30.36 ± 14.23	30.70 ± 13.45

a well-defined mathematical smoothing formulation is applied to ensure a stable and consistent FBI estimate.

The FBI estimates presented in Table 2 are based on the extensive NFHS-IV dataset. Recognizing the challenges associated with consistently collecting such extensive data is crucial. Therefore, for individual studies, those with budgetary constraints, or those with shorter duration requirements, simulation methods are pursued. These simulation techniques assist in determining the appropriate sample size required for the effective implementation of the current status data technique.

Result based on Simulation Study for finding Minimum Sample Size

As shown in Table 1, for sample sizes of 1500, the length of confidence intervals for the mean using Spline Smoothing method is 10.01. However, when Ayer Method is used, as shown in Table 1, the length of the confidence interval for the mean is 4.15. For a sample size of 5000, the length of confidence intervals for the mean is 4.76 with Spline Smoothing method. On the other hand, when employing the Ayer Method, the required sample size to 4.76 length of confidence interval is 1500.

Additionally, when applying the Ayer method, a noteworthy observation is made. As the sample size increases, the confidence interval for FBI estimates narrows considerably when compared to spline smoothing method. This narrowing of the confidence interval is primarily attributed to the reduction in the standard error of the FBI estimate achieved through the current status method. However, it is important to note that the FBI estimate is insignificant (till sample size reaches 2500 and 3000) due to slight bias. This highlight the method's robustness and precision with larger datasets (see figure S1 and S2 in supplementary).

Sample Size estimation using Ayer Method

Despite the simulation technique, the mathematical formula for sample size is shown by Equation 2. On considering different value of margin of error and η, get different sample size (as shown in Figure 2). On considering at 20% of margin on error and 10% of η the

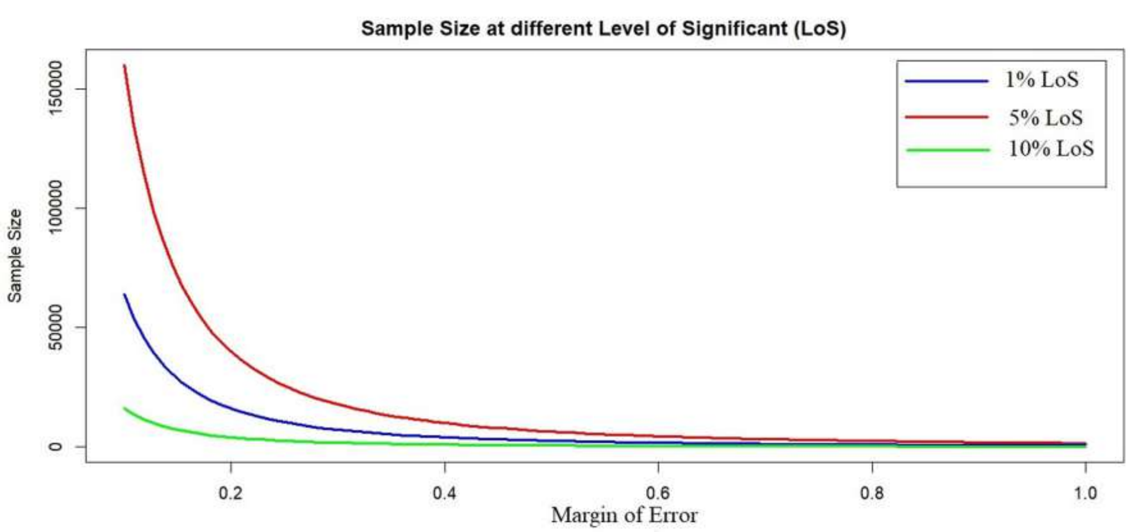

Figure 2. Sample size at different level of significant with Margin of Error

Table 3. Sample Size at different level of Significance (η) and Margin of Error (ϵ)

S.no.	Margin of Error (ϵ)	Sample size at 1% η	Sample size at 5% η	Sample size at 10% η
1	0.1	160000.00	32000.00	16000.00
2	0.10	134444.40	26888.89	13444.44
3	0.11	114556.20	22911.24	11455.62
4	0.2	40000.00	8000.00	4000.00
5	0.3	17777.78	3555.55	1777.77
6	0.4	10000.00	2000.00	1000.00
7	0.5	6400.00	1280.00	640.00
8	0.6	4444.44	888.88	444.44
9	0.7	3265.30	653.06	326.53
10	0.8	2500.00	500.00	250.00
11	0.9	1975.30	395.06	197.53
12	1	1600.00	320.00	160.00

sample size is 2500 which is approximate equal to simulation result of Table 3.

Discussion

The analysis of first birth intervals, particularly when using cross-sectional or interval-censored data, poses significant challenges due to truncation. Truncation occurs when individuals enter the study at various times after marriage, resulting in incomplete observation of their first birth intervals. This can introduce bias if not adequately adjusted, leading to either overestimation or underestimation of the actual birth interval distribution. In this context, we have employed the current status data technique to address truncation effects. By utilizing the Ayer method, a nonparametric maximum likelihood estimator, we can estimate the distribution of first birth intervals while accounting for the censored nature of the data and the varying entry points of the individuals in the study.

Compared to traditional methods, such as the life table or parametric models, the current status technique provides a more robust and flexible solution. Keiding (1991),[33] for example, employed parametric models to estimate age-specific fertility rates but did not fully account for the truncation observed in cross-sectional data. In contrast, the Ayer method avoids strong distributional assumptions, making it more applicable in cases where the underlying data distribution is not well understood. While this method is effective, it relies on categorizing time into intervals, potentially reducing precision. Our nonparametric approach, on the other hand, maintains continuous-time estimates and offers more accurate adjustments for truncation.

Other researchers, such as Kalbfleisch and Prentice (1980),[23] applied semi-parametric proportional hazards models to time-to-event data. Although their models can handle truncation by incorporating covariates, they require the assumption of proportional hazards, which may not always be applicable in fertility data. The Ayer method, however, offers greater flexibility by avoiding such assumptions, making it particularly useful when the proportional hazards assumption does not hold. Furthermore, Diamond et al. (1986)

[26] used parametric models like the log-normal distribution to adjust for truncation in first birth intervals. While parametric approaches are valuable when the underlying distribution is known, they can lead to biased results if the model is misspecified. In contrast, our use of the Ayer method is more robust, as it does not require prior knowledge of the distribution.

Overall, the current status technique has proven to be an effective method for adjusting truncation effects in the analysis of first birth intervals. It provides unbiased estimates without relying on assumptions about the data's underlying distribution, making it more versatile in demographic studies where truncation and censoring are prevalent. Traditional methods like life tables and parametric models have their advantages. However, the current status approach offers more flexibility and robustness, especially when event times are only partially observed or the population shows variation in timing and covariate effects. Future research can extend this approach by incorporating covariate adjustments and handling more complex truncation scenarios, thereby broadening its applicability across different fields of demographic and health research.

Conclusion

The study explored different methods and sample sizes for estimating the First Birth Interval (FBI), focusing on the impact of truncation effects and data limitations. By comparing various approaches, the research highlights the complexities involved in accurately estimating FBI, particularly when dealing with incomplete or truncated data. The findings underscore the importance of considering factors like censoring, selection bias, and truncation in demographic studies. Additionally, the study provides a framework for selecting appropriate methods and determining sample sizes in cross-sectional surveys, ultimately offering more reliable estimates of FBI, especially in cases where traditional longitudinal data is unavailable. These insights are significant for researchers and policymakers aiming to understand birth patterns and reproductive behaviors in diverse populations.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

1. M Mazharul Al Shanfari, Noora Said Salim & Islam. Survival analysis and predictors of time to first birth after marriage in jordan. International Journal of Statistics and Applications, 11(3):51–60, 2021.

2. Nancy Brandon Tuma and Michael T Hannan. Approaches to the censoring problem in analysis of event histories. Sociological methodology, 10:209–240, 1979.

3. Kwan-Moon Leung, Robert M Elashoff, and Abdelmonem A Afifi. Censoring issues in survival analysis.Annual review of public health, 18(1):83–104, 1997.

4. Rahmat Hidayat, Hadi Sumarno, Endar H Nugrahani, et al. Survival analysis in modeling the birth interval of the first child in indonesia. Open Journal of Statistics, 4(03):198, 2014.

5. Anup Kumar, RC Yadava, and U Srivastava. Analysing the impact of marital duration on first birth interval. Journal of Statistics and Applications, 5(3/4):195, 2010.

6. Mindel C Sheps, Mindel S Sheps, Jane A Menken, and Annette P Radick. Mathematical models of concep- tion and birth. University of Chicago Press, 1973.

7. W Kenneth Poole. Fertility measures based on birth interval data. Theoretical Population Biology, 4(3): 357–387, 1973.

8. Mindel C Sheps and Jane A Menken. Distribution of birth intervals according to the sampling frame. Theo- retical Population Biology, 3(1):1–26, 1972.

9. SN Singh, BN Bhattacharya, and RC Yadava. An adjustment of a selection bias in postpartum amenorrhea from follow-up studies. Journal of the American Statistical Association, 74(368):916–920, 1979.

10. Rolf Ulrich and Jeff Miller. Effects of truncation on reaction time analysis. Journal of Experimental Psychology: General, 123(1):34, 1994.

11. David Wolfers. Determinants of birth intervals and their means. Population Studies, 22(2):253–262, 1968.

12. Anup Kumar and RC Yadava. Usual closed birth interval versus most recent closed birth interval. Journal of Data Science, 13(1):73–93, 2015.

13. Joseph Molitoris. Heterogeneous effects of birth spacing on neonatal mortality risks in bangladesh. Studies in Family Planning, 49(1):3–21, 2018.

14. Amusan Ajitoni Simeon, Zarina Mohd Khalid, and Bahru Malaysia. Survival modeling of first birth interval after marriage. Life Sci J, 11:11–14, 2014.

15. Fan Yang and Peng Ding. Using survival information in truncation by death problems without the mono- tonicity assumption. Biometrics, 74(4):1232–1239, 2018.

16. SS Sharma. A study on some mathematical models for birth intervals. PhD thesis, PhD thesis, Banaras Hindu University, India, 2004.

17. Anup Kumar, Abhishek Bharti, and RC Yadava. An adjustment of truncation and selection effect for esti- mating conception rate from first birth interval data. Journal of Data Science, 17(4):756–767, 2019.

18. Anup Kumar and RC Yadava. Analysing censored data on first birth interval for heterogeneous group of females. J. Comb. Inf. Syst. Sci., 39(4):149–165, 2014.

19. Wei-Yann Tsai and Cun-Hui Zhang. Asymptotic properties of nonparametric maximum likelihood estimator for interval-truncated data. Scandinavian journal of statistics, pages 361–370, 1995.

20. Niels Keiding, Oluf K HøJbjerg Hasen, Ditte Nørbo Sørensen, and Rémy Slama. The current duration approach to estimating time to pregnancy. Scandinavian Journal of Statistics, 39(2):185–204, 2012.

21. Chunjie Wang, Bo Zhao, Linlin Luo, and Xinyuan Song. Regression analysis of current status data with latent variables. Lifetime Data Analysis, 27(3):413–436, 2021.

22. Anup Kumar, Sachin Kumar, Jai Kishun, Uttam Singh, and Pushpraj Pushpraj. Estimation of the distribution of duration of breastfeeding from cross-sectional data: Some methodological issue. Journal of Biostatistics and Epidemiology, 6(4):290–304, 2020.

23. Jianguo Sun and John D Kalbfleisch. The analysis of current status data on point processes. Journal of the American Statistical Association, 88(424):1449–1454, 1993.

24. Chunjie Wang, Jianguo Sun, Liuquan Sun, Jie Zhou, and Dehui Wang. Nonparametric estimation of current status data with dependent censoring. Lifetime data analysis, 18(4):434–445, 2012.

25. Mindel C Sheps, Jane A Menken, Jeanne Clare Ridley, and Joan W Lingner. Truncation effect in closed and open birth interval data. Journal of the American Statistical Association, 65(330):678–693, 1970.

26. Ian D Diamond, John W McDonald, and Iqbal H Shah. Proportional hazards models for current status data: application to the study of differentials in age at weaning in pakistan. Demography, pages 607–620, 1986.

27. David A Schoenfeld. Sample-size formula for the proportional-hazards regression model. Biometrics, pages 499–503, 1983.

28. Alan B Cantor. Sample size calculations for the log rank test: a gompertz model approach. Journal of clinical epidemiology, 45(10):1131–1136, 1992.

29. John M Williamson, Hung-Mo Lin, and Hae-Young Kim. Power and sample size calculations for current status survival analysis. Statistics in medicine, 28(15):1999–2011, 2009.

30. Ian C Marschner. Determining the size of a cross-sectional sample to estimate the age-specific incidence of an irreversible disease. Statistics in medicine, 13(22):2369–2381, 1994.

31. DY Lin, David Oakes, and Zhiliang Ying. Additive hazards regression with current status data. Biometrika, 85(2):289–298, 1998.

32. Janice M Morse. Determining sample size, 2000.

33. Niels Keiding. Age-specific incidence and prevalence: a statistical perspective. Journal of the Royal Statistical Society Series A: Statistics in Society, 154(3):371–396, 1991.

34. A Meredith John, Jane A Menken, and James Trussell. Estimating the distribution of interval length: Current status and retrospective history data. Population Studies, 42(1):115–127, 1988.

35. Cosma Shalizi. Advanced data analysis from an elementary point of view. pages 177–192, 2013.

36. Miriam Ayer, H Daniel Brunk, George M Ewing, William T Reid, and Edward Silverman. An empirical distribution function for sampling

with incomplete information. The annals of mathematical statistics, pages 641–647, 1955.

37. International Institute for Population Sciences (IIPS) and ICF. National family health survey (nfhs-4), india, 2015-16. Data, 2017. https://dhsprogram.com/data/available-datasets.cfm.

Appendixes

Table 1. Distribution function of Current Status Data for (First Birth Interval) FBI ti ≤ 120 months

t_j	y_j	$(n_j - y_j)$	n_j	Spline Smoothing		Ayer Method	
				Unsmoothed	Smoothed	S_j	
0	963	0	963	1	1	1	1
1	2107	4	2111	1	1	1	1
2	1968	3	1971	1	1	1	1
3	1738	5	1743	1	1	1	1
4	1433	2	1435	1	0.99	1	1
5	1206	2	1208	1	0.97	1	1
6	1055	3	1058	1	0.94	1	1
7	919	1	920	1	0.91	1	1
8	953	1	954	1	0.88	1	1
9	1349	27	1376	0.98	0.85	0.98	0.98
10	1453	110	1563	0.93	0.82	0.93	0.93
11	1738	226	1964	0.88	0.79	0.88	0.88
12	1964	450	2414	0.81	0.76	0.81	0.81
13	1964	555	2519	0.78	0.73	0.78	0.78
14	1495	601	2096	0.71	0.7	0.71	0.71
15	1054	601	1655	0.64	0.67	0.64	0.64
16	820	571	1391	0.59	0.64	0.59	0.59
17	573	463	1036	0.55	0.61	0.55	0.55
18	419	445	864	0.48	0.58	0.48	0.48
19	407	467	874	0.47	0.55	0.47	0.47
20	417	502	919	0.45	0.52	0.45	0.46
21	627	723	1350	0.46	0.5	0.46	0.46
22	670	920	1590	0.42	0.47	0.42	0.42
23	804	1084	1888	0.43	0.45	0.43	0.42
24	896	1320	2216	0.4	0.42	0.4	0.4
25	849	1359	2208	0.38	0.4	0.38	0.38
26	770	1248	2018	0.38	0.38	0.38	0.38
27	543	1100	1643	0.33	0.36	0.33	0.33
28	398	924	1322	0.3	0.34	0.3	0.3
29	319	720	1039	0.31	0.32	0.31	0.3
30	253	656	909	0.28	0.31	0.28	0.28
31	219	632	851	0.26	0.29	0.26	0.27

t_j	y_j	$(n_j - y_j)$	n_j	Spline Smoothing		Ayer Method	
				Unsmoothed	Smoothed	S_j	
32	246	685	931	0.26	0.28	0.26	0.27
33	344	932	1276	0.27	0.27	0.27	0.27
34	449	1225	1674	0.27	0.25	0.27	0.27
35	511	1460	1971	0.26	0.24	0.26	0.26
36	558	1814	2372	0.24	0.23	0.24	0.25
37	596	1700	2296	0.26	0.22	0.26	0.25
38	499	1546	2045	0.24	0.21	0.24	0.24
39	364	1347	1711	0.21	0.2	0.21	0.21
40	275	1123	1398	0.2	0.2	0.2	0.2
41	201	880	1081	0.19	0.19	0.19	0.19
42	133	750	883	0.15	0.18	0.15	0.17
43	132	717	849	0.16	0.17	0.16	0.17
44	152	745	897	0.17	0.17	0.17	0.17
45	213	1008	1221	0.17	0.16	0.17	0.17
46	301	1325	1626	0.19	0.16	0.19	0.17
47	322	1626	1948	0.17	0.15	0.17	0.17
48	365	1924	2289	0.16	0.15	0.16	0.16
49	348	2000	2348	0.15	0.15	0.15	0.15
50	282	1675	1957	0.14	0.14	0.14	0.14
51	206	1372	1578	0.13	0.14	0.13	0.13
52	172	1173	1345	0.13	0.13	0.13	0.13
53	119	937	1056	0.11	0.13	0.11	0.13
54	114	705	819	0.14	0.13	0.14	0.13
55	113	796	909	0.12	0.13	0.12	0.13
56	129	821	950	0.14	0.12	0.14	0.13
57	180	1153	1333	0.14	0.12	0.14	0.13
58	225	1545	1770	0.13	0.12	0.13	0.13
59	244	1839	2083	0.12	0.12	0.12	0.12
60	297	2082	2379	0.12	0.12	0.12	0.12
61	282	2114	2396	0.12	0.11	0.12	0.12
62	225	1703	1928	0.12	0.11	0.12	0.12
63	188	1383	1571	0.12	0.11	0.12	0.12
64	127	1202	1329	0.1	0.11	0.1	0.11
65	131	860	991	0.13	0.11	0.13	0.11
66	77	789	866	0.09	0.11	0.09	0.11
67	85	732	817	0.1	0.1	0.1	0.11
68	103	900	1003	0.1	0.1	0.1	0.11
69	166	1262	1428	0.12	0.1	0.12	0.11
70	195	1525	1720	0.11	0.1	0.11	0.11
71	207	1852	2059	0.1	0.1	0.1	0.1

t_j	y_j	$(n_j - y_j)$	n_j	Spline Smoothing		Ayer Method	
				Unsmoothed	Smoothed	S_j	
72	229	1966	2195	0.1	0.1	0.1	0.1
73	207	1974	2181	0.09	0.1	0.09	0.09
74	167	1703	1870	0.09	0.09	0.09	0.09
75	130	1436	1566	0.08	0.09	0.08	0.09
76	100	1137	1237	0.08	0.09	0.08	0.09
77	82	852	934	0.09	0.09	0.09	0.09
78	84	711	795	0.11	0.09	0.11	0.09
79	59	651	710	0.08	0.09	0.08	0.09
80	70	749	819	0.09	0.09	0.09	0.09
81	125	1037	1162	0.11	0.09	0.11	0.09
82	125	1374	1499	0.08	0.08	0.08	0.08
83	145	1666	1811	0.08	0.08	0.08	0.08
84	182	1965	2147	0.08	0.08	0.08	0.08
85	153	1880	2033	0.08	0.08	0.08	0.08
86	152	1701	1853	0.08	0.08	0.08	0.08
87	123	1424	1547	0.08	0.08	0.08	0.08
88	85	1151	1236	0.07	0.08	0.07	0.08
89	72	844	916	0.08	0.08	0.08	0.08
90	63	741	804	0.08	0.08	0.08	0.08
91	47	686	733	0.06	0.08	0.06	0.08
92	61	764	825	0.07	0.08	0.07	0.08
93	89	1014	1103	0.08	0.08	0.08	0.08
94	116	1438	1554	0.07	0.08	0.07	0.08
95	133	1766	1899	0.07	0.07	0.07	0.08
96	165	2003	2168	0.08	0.07	0.08	0.08
97	161	1988	2149	0.07	0.07	0.07	0.08
98	151	1768	1919	0.08	0.07	0.08	0.08
99	106	1393	1499	0.07	0.07	0.07	0.08
100	84	1076	1160	0.07	0.07	0.07	0.08
101	59	818	877	0.07	0.07	0.07	0.08
102	55	608	663	0.08	0.07	0.08	0.08
103	53	627	680	0.08	0.07	0.08	0.08
104	61	762	823	0.07	0.07	0.07	0.08
105	103	1033	1136	0.09	0.07	0.09	0.08
106	90	1489	1579	0.06	0.07	0.06	0.07
107	125	1674	1799	0.07	0.07	0.07	0.07
108	148	1970	2118	0.07	0.07	0.07	0.07
109	126	1817	1943	0.06	0.07	0.06	0.07
110	120	1572	1692	0.07	0.07	0.07	0.07
111	93	1279	1372	0.07	0.07	0.07	0.07

t_j	y_j	$(n_j - y_j)$	n_j	Spline Smoothing		Ayer Method	
				Unsmoothed	Smoothed	S_j	
112	80	1019	1099	0.07	0.07	0.07	0.07
113	47	797	844	0.06	0.07	0.06	0.07
114	49	639	688	0.07	0.07	0.07	0.07
115	42	601	643	0.07	0.07	0.07	0.07
116	66	772	838	0.08	0.07	0.08	0.07
117	92	1106	1198	0.08	0.07	0.08	0.07
118	106	1490	1596	0.07	0.07	0.07	0.07
119	132	1701	1833	0.07	0.07	0.07	0.07
120	151	2097	2248	0.07	0.07	0.07	0.07

Where

T_i is the time

be the number of women having marital duration in the interval months at the time of survey,

be the number of women with martial duration in the interval months who don't have the first birth (i.e., the FBI is greater than months) at the time of survey, j = 1, 2, . . . , k.

$S(tj) = \frac{y_j}{n_j} =$ "proportion of women whose FBI is greater than " t_j, j=1,2,…,k.

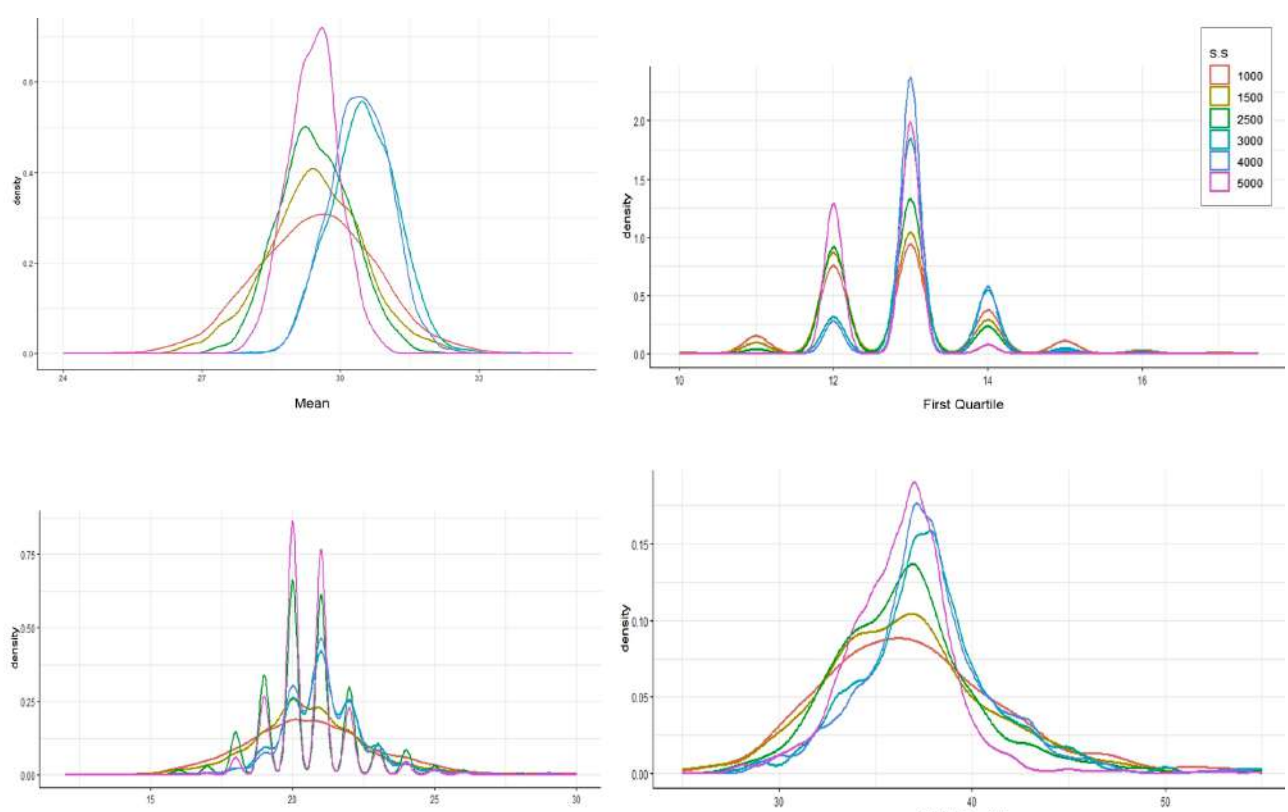

Figure 1. Distribution of Statistics with different Sample Size by Ayer Method

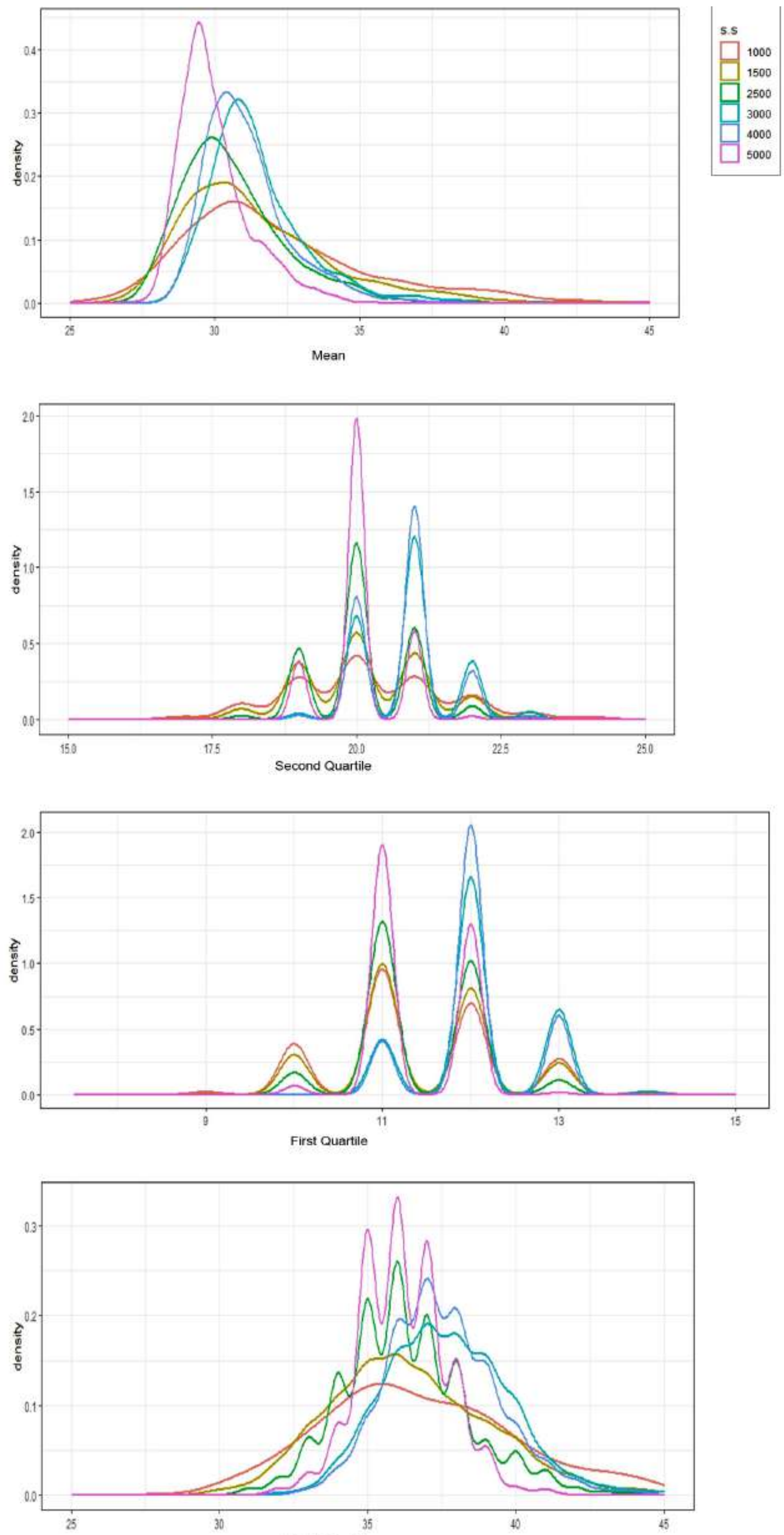

Figure 2. Distribution of Statistics with different Sample Size by Spline smoothing