Journal of Data Science,17(4). P. 756 -767,2019
DOI:10.6339/JDS.201910_17(4).0008

An Adjustment of Truncation and Selection Effect for Estimating Conception Rate from First Birth Interval Data

Anup Kumar[1], Abhishek Bharti[2], R. C. Yadava[2]

[1] *Department of Biostatistics and Health Informatics, SGPGIMS*

[2] *Department of Statistics, Institute of Science, Banaras Hindu University*

ABSTRACT

A technique is proposed to estimate the conception rate using the distribution of first birth interval of recently married women. The proposed technique adjusts the truncation and selection effects present in a cross-sectional data. Real data from NFHS-3 and NFHS-4 are used for illustration.

Key words: Truncation Bias, Selection effect, Conception Rate, Heterogeneity.

1. Introduction

The study of human fertility through birth interval approach is very popular among demographers and statisticians. The data of birth interval are considered as an indicator of reproductive performance and difference in observed distributions reflects the different patterns of reproduction (see Yadava et al. (2013)). A survey includes varying age and marital duration due to which the study of the birth interval as a fertility measure has suered from truncation, censoring and selectivity.

In a cross-sectional survey, married women have different marital durations. Female having marital duration T can never have the value of first conception time X greater than T. Hence, for a shorter T, the observed values of X will be small since many women are subject to truncation as they did not give a birth in (0, T). To avoid truncation bias, analysis of first birth interval was done only with the woman of higher marital duration.

Generally, for mathematical modelling and other practical purposes characteristics of the population are assumed as homogeneous in nature, but, in reality, it is of heterogeneous nature. Heterogeneity in the underlying population places diculties in the way of interpretation of all statistical data based on averages.

A type of selectivity occurs in the heterogeneous population. Those women who are more fertile will conceive sooner and drop-outs for observation will be those who are less fertile. In a follow-up study, if we take a cohort of married woman and follow the individuals in it month by month to find the duration of first birth, we may face the selection effect of heterogeneity. Wolfers (1968) discussed the problem associated with heterogeneity among women with respect to their fertility parameters. Singh et al. (1979) also tackle the problem associated with heterogeneity which has been termed as Selection Bias in the analysis of duration of post-partum amenorrhea.

Theoretically, truncation bias and selection bias can be handled through the proper technique of standardization. Sheps et al. (1970) have discussed the issue of truncation effect. Later on, Sheps and Menken (1972) and Sheps and Menken (1973) added another dimension known as the sampling frame. Besides truncation effect, heterogeneity with respect to parameters involved in a model sampling frame may also alter the distribution significantly. This typically describes the difficulties of the impact of sampling frame on the distribution of birth intervals. Sampling frame (ascertainment plan) for a birth interval refers to the manner in which a birth interval is determined. Apart from other considerations, the ascertainment plan or sampling frame influences the distribution of birth interval so much so that many times one may draw incorrect or misleading conclusions (see Kumar and Yadava (2015)).

Development of a model for the duration of first conceptive time has been possible under the assumption that conception depends on chance. Time from marriage to conception and then birth has been treated as either discrete or continuous in the literature. Treating fecundability as constant till the woman conceives and time as a discrete random variable, the waiting time for the first conception has been described by geometric distribution (see Gini (1924); Henry (1961)). Potter and Parker (1964) and Sheps (1964) generalized the above expressions by incorporating heterogeneity and the chance of foetal losses before the first live births. Models using negative exponential distribution have been advanced to describe the above situations when the time is treated as continuous. In the case of heterogeneous population, it is generally assumed that the parameter involved follows a type III Pearson distribution Singh (1964). Researchers have also argued that the fecundability changes over time. Bhattacharya et al. (1988) derived a model for the time of first birth under the assumption that the exposure to the risk of conception is delayed due to the visit of the woman to her parent's house and hence the fecundability is less in the beginning and as age advances it reaches the maximum.

To overcome the impact of truncation bias many authors have analyzed only those women whose marital duration is large enough to have probability one for having first birth. But this frame has some major drawbacks. The estimate obtained from women of larger marital duration is not current and the chance of recall lapse is higher among women having large marital duration consequently they may not report the duration of first birth interval precisely. Further, the sample size is small as it excludes a considerable number of recently married women. Kumar et al. (2010) have analyzed the behavior of the distribution of the first conception in shorter marital duration accounting to the homogeneous group of woman.

In the next section, a procedure is described to estimate fecundability from the first birth interval in the shorter marital duration owing to a heterogeneous population. Certain conditions and assumptions are considered for the formulation of these theoretical results. In section 3, how the distribution of the first conception behaves in the shorter marital duration is investigated. In section 4, real data from NFHS-3 (2007) and NFHS-4 (2016) are used for illustration.

2. The Procedure

A probability model is proposed for the waiting time to the first conception including a heterogeneous group of women of shorter marital duration. Let $X > 0$ be writing time for first conception and $f(x)$ probability density function (pdf) and $F(x)$ be the cumulative distribution function (cdf) of X. If X is truncated at T, then the truncated mean of X will be

$$E_1(T) = \frac{\int_0^T xf(x)dx}{F(T)} \tag{1}$$

Here, T may be small or large. In a retrospective survey, women may have varying marital durations and the exposure time for conception will also be varying among women. Consequently, T (exposure time) may be a assumed as a random variable with pdf $\xi(t)$. Thus, the mean waiting time to the first conception for the woman having exposure time $(0, T)$ is given as

$$E_2(T) = \frac{\int_0^T E_1(t)\xi(t)dx}{\int_0^T \xi(t)dx} \tag{2}$$

The Eqs. 1 and 2 are general in nature and values of $E_1(X)$ and $E_2(X)$ can be obtained by specifying the density $f(x)$ and $\xi(t)$. The distribution of X for heterogeneous groups of women with respect to λ has been commonly used by various authors earlier also Singh (1964). Chakraborty (1976), Pratap (2011), Kumar (2012) and other as described in next section.

2.1 Heterogeneous Conception Rate

If the conception rate λ of women is assumed to be constant for all women then waiting time to the first conception time follows the exponential distribution. Let us consider that λ varies among woman and, it may be treated as a random variable following a specific distribution. Under these conditions, the pdf of X can be derived as

$$f(x) = \int_0^\infty \lambda e^{-\lambda x} \mathrm{g}(\lambda) d\lambda \tag{3}$$

where $\mathrm{g}(\lambda)$ is the p.d.f. of the conception rate λ.

This mixture can take a parametric form or be left arbitrary. The most widely published model for heterogeneity in conception rate assumes Gamma distribution for λ.

$$\mathrm{g}(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-\lambda b}; a > 0, b > 0 \tag{4}$$

The choice of the distribution is due to its flexibility, mathematical applicability and interpretation. Under this situation using equation 3, $f(x)$ becomes

$$f(x) = \frac{ab^a}{(x+b)^{a+1}};\ a > 0, b > 0, x > 0 \tag{5}$$

Using equation 1 truncated mean at T is given by

$$E_1(T) = \frac{\int_0^T x \frac{ab^a}{(x+b)^{a+1}} dx}{\int_0^T \frac{ab^a}{(x+b)^{a+1}} dx} = \frac{ab^a \left\{ \left[\frac{(b+T)^{1-a} - b^{1-a}}{1-a} \right] + b \left[\frac{(b+T)^{-a} - b^{-a}}{a} \right] \right\}}{1 - (\frac{b}{b+T})^a} \tag{6}$$

In order to find the value of $E_2(X)$, we have to specify the distribution of $\xi(t)$. If it is assumed that the population is stationary and marriage pattern does not change over time then the number of marriages will be uniform over time. This assumption is quite reasonable for smaller value of T i.e. $T \le 10$ year. Consequently it is assumed that $\xi(t)$ follow uniform distribution. Hence, from equation 2 the expression for $E_2(X)$, under the assumption that

$$\xi(t) = 1/T\ ; 0 < t < T$$

is given as ($E_2(T)$ may be termed as combined mean)

$$E_2(T) = \frac{1}{T} \int_0^T E_1(T) dt \tag{7}$$

3. Simulation Study

To simulate the two derived expressions 6 and 7, R statistical software has been used. The value of $E_1(X)$ and $E_2(X)$ at different values of T and of a and b are given in Tables 1 & 2; and plotted in Figures 1 & 2. The value of a and b are taken from Kumar and Yadava (2014), which are estimated values from a real date set of NFHS-3 (2007).

Table 1: Mean Waiting Time to First Conception (in months) i.e., $E_1(T)$ for different values of T (in years) and $\boldsymbol{a}$ & $\boldsymbol{b}$

T	Combinations of a and b				
	a=2.62 b=32.46	a=2.47 b=30.46	a=2.33 b=28.56	a=2.27 b=27.81	a=2.16 b=26.36
1	4.876	4.862	4.848	4.842	4.830
2	8.146	8.121	8.096	8.085	8.063
3	10.437	10.415	10.392	10.383	10.364
4	12.105	12.096	12.086	12.083	12.076
5	13.357	13.368	13.379	13.386	13.396
6	14.322	14.357	14.393	14.410	14.443
7	15.083	15.144	15.205	15.235	15.290
8	15.694	15.781	15.869	15.911	15.990
9	16.193	16.306	16.420	16.474	16.577
10	16.606	16.743	16.884	16.950	17.075
11	16.952	17.113	17.278	17.356	17.504
12	17.246	17.429	17.617	17.706	17.875
13	17.497	17.702	17.912	18.011	18.201
14	17.713	17.938	18.170	18.278	18.488
15	17.902	18.145	18.397	18.515	18.743
16	18.067	18.328	18.598	18.725	18.970
17	18.212	18.490	18.778	18.913	19.175
18	18.341	18.634	18.939	19.082	19.360
19	18.456	18.763	19.084	19.235	19.528
20	18.558	18.880	19.215	19.373	19.681
21	18.651	18.985	19.334	19.499	19.820
22	18.734	19.081	19.443	19.615	19.949
23	18.810	19.168	19.543	19.721	20.067
24	18.879	19.248	19.635	19.818	20.176
25	18.941	19.321	19.719	19.908	20.277
26	18.999	19.388	19.797	19.991	20.371
27	19.052	19.450	19.869	20.068	20.458
28	19.100	19.507	19.936	20.140	20.540
29	19.145	19.560	19.998	20.207	20.616
30	19.186	19.609	20.056	20.269	20.687

Table 2: Combine Mean Waiting Time to First Conception (in months) i.e. $E_2(T)$ for different values of T (in years) and $\boldsymbol{a}$ & $\boldsymbol{b}$

T	Combinations of a and b				
	a=2.62 b=32.46	a=2.47 b=30.46	a=2.33 b=28.56	a=2.27 b=27.81	a=2.16 b=26.36
1	2.609	2.603	2.597	2.595	2.590
2	4.611	4.598	4.583	4.579	4.568
3	6.192	6.176	6.156	6.151	6.136
4	7.472	7.456	7.435	7.432	7.417
5	8.530	8.517	8.498	8.498	8.486
6	9.418	9.411	9.397	9.401	9.395
7	10.175	10.176	10.169	10.177	10.179
8	10.828	10.838	10.840	10.853	10.863
9	11.397	11.417	11.428	11.447	11.466
10	11.898	11.929	11.949	11.975	12.003
11	12.342	12.384	12.415	12.446	12.484
12	12.739	12.791	12.833	12.870	12.918
13	13.096	13.159	13.211	13.254	13.312
14	13.418	13.492	13.555	13.603	13.672
15	13.711	13.795	13.869	13.923	14.001
16	13.978	14.073	14.157	14.217	14.305
17	14.223	14.328	14.422	14.488	14.586
18	14.448	14.563	14.667	14.738	14.846
19	14.656	14.781	14.895	14.971	15.088
20	14.849	14.983	15.107	15.188	15.314
21	15.027	15.171	15.304	15.390	15.525
22	15.194	15.347	15.489	15.580	15.723
23	15.350	15.511	15.662	15.757	15.910
24	15.495	15.665	15.824	15.924	16.085
25	15.632	15.810	15.977	16.082	16.251
26	15.760	15.946	16.122	16.231	16.407
27	15.881	16.075	16.258	16.372	16.556
28	15.995	16.196	16.388	16.505	16.697
29	16.103	16.312	16.510	16.631	16.831
30	16.205	16.421	16.627	16.752	16.958

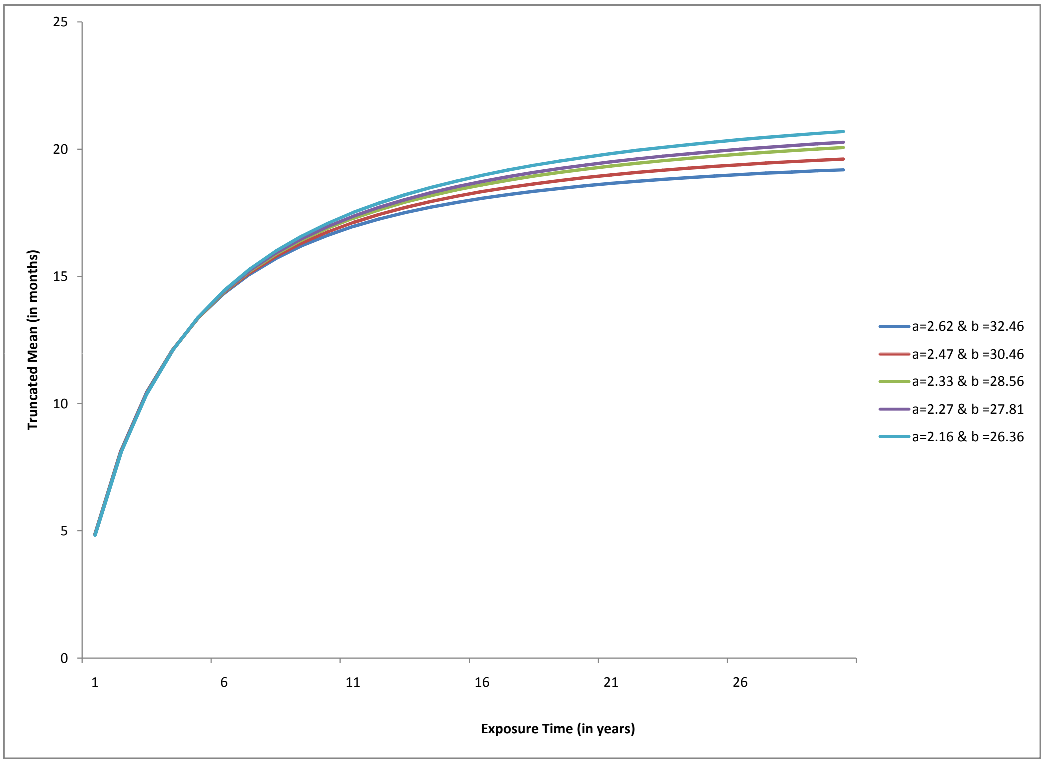

Figure 1: Truncated mean $E_1(T)$ for different values of $\boldsymbol{a}, \boldsymbol{b}$ and T

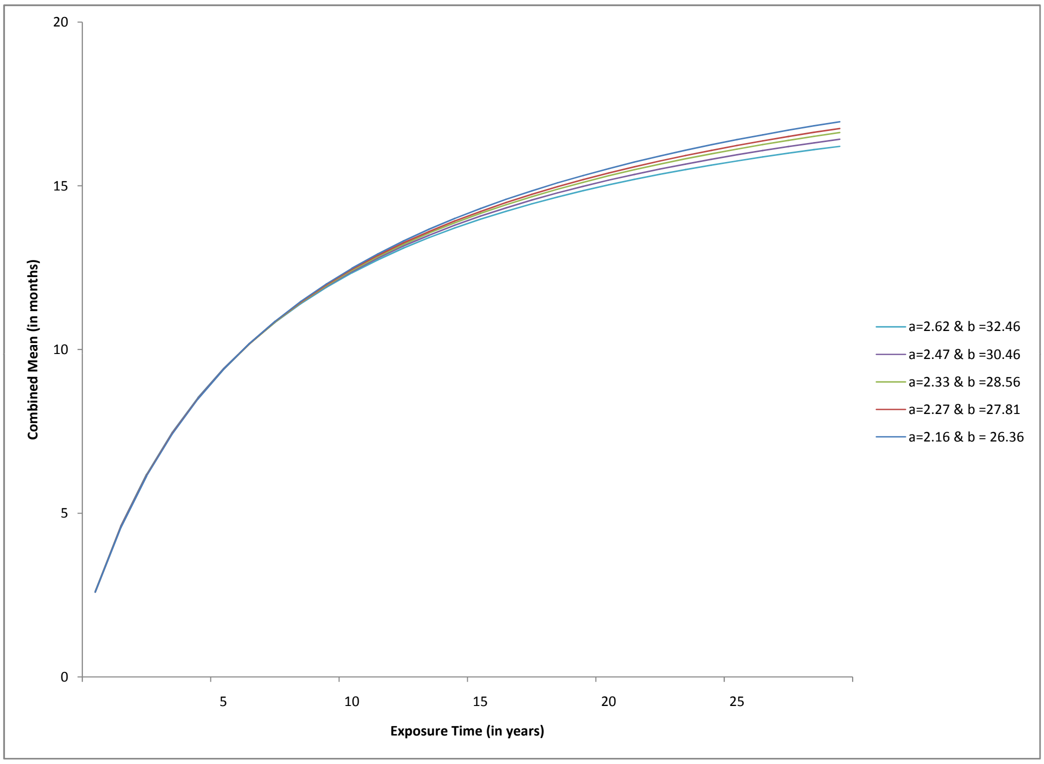

Figure 2: Combined mean $\boldsymbol{E_2(T)}$ for different values of $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{T}$

Table 1 and figure 1 illustrate the truncated mean $E_1(T)$ is significantly affected by the shorter value of $\boldsymbol{T}$, but this impact becomes small for large values of $\boldsymbol{T}$. For a=2.62 and $b = 32.46$, the mean conception time $\boldsymbol{E_1(T)}$ show little variation after $\boldsymbol{T} = 16$ whereas show no variation after $\boldsymbol{T} = 27$. Somehow similar trend is also observed in Table 2 and Figure 2 for combined mean. For a=2.62 and $b = 32.46$, the combine mean conception time $\boldsymbol{E_2(T)}$does not show any variation after $\boldsymbol{T} = 29$. These simulation results encourage to evaluate the parameters of proposed model from real data.

4. Application

The proposed methodologies are applied to data obtained from for a National Family Health Survey; NFHS-3 (2005 - 2006) and NFHS-4 (2015 - 2016). Data on age at marriage, date of marriage, dates of first birth interval and date of the survey is available. Using this, marital duration and first birth interval of each woman is obtained. It is assumed that the gestation period is of 9 months for each woman. Hence, duration of the first conception is estimated by subtracting 9 months from the duration of the first birth interval. For sake of simplicity, many authors consider assumption of one to one correspondence between conception and birth. The woman of lower age at marriage may have different fertility behavior before first birth than the woman of higher age at marriage. Hence, only the woman whom age at marriage is 16 years and above are considered.

There are two unknown parameters a and b in the model. The expression for $E_2(T)$is implicit in nature so for finding the value of $E_2(T)$, numerical integration is applied against specified values of a, b and T. To find the ML estimate for a and b, Gauss-Legendre quadrature is used. Even with the implementation of the above-said procedure, reliable estimates of a and b is not obtained with no certainty regarding the existence of local or global maxima. This may be due to the convergence problem of the implicit expression of $E_2(T)$ for certain values of a, b. Hence, for estimating a and b, we equate observed mean and variance to theoretical mean and variance. In order to check the validity of the principle of moments, we simulate the value of $E_2(T)$ against specied values of a and b derived from the mean and variance equation. The range of guess values for a and b was obtained from earlier studies. we get a unique value for a and b for the specied value of T by equating observed mean and variance to theoretical mean and variance.

The values of time of the first conception (in months) for different values of T considering woman with exposure time <= T years for $T = 5,6,7,8,9$ and 10 years are presented in Table 3 for NFHS-3 and 4 for NFHS-4 along with the estimate of of $\hat{a}$, $\hat{b}$ heterogeneous $\hat{\lambda}$ (column 5) given as

$$\hat{\lambda} = \frac{\hat{a}}{\hat{b}}, \tag{8}$$

and Harmonic mean of Conception Rate $\hat{\lambda}_H$ (column 6) is given as

$$\hat{\lambda}_H = \frac{\hat{a} - 1}{\hat{b}} \tag{9}$$

Methodology for yearly conception rate by Kumar et al. (2010) is also used to estimate the (yearly conception rate) for the homogeneous group of women presented in column 3 of Table 4 and 5.

Table 3: Estimated value of a, b and λ at different values of T for women of ages at marriage 16 year and above from NFHS 3 data

T	Mean Conception	Homogeneous $\hat{\boldsymbol{\lambda}}$	$\hat{\boldsymbol{a}}$	$\hat{\boldsymbol{b}}$	Heterogeneous $\hat{\boldsymbol{\lambda}}$	$\hat{\boldsymbol{\lambda}}_{\boldsymbol{H}}$
5	9.3468	0.0688	2.34	36.40	0.0642	0.0368
6	10.1756	0.0675	2.30	33.66	0.0683	0.0386
7	10.6587	0.0675	2.31	31.12	0.0742	0.0420
8	11.0946	0.0679	2.82	36.79	0.0766	0.0494
9	11.4513	0.0680	2.83	35.69	0.0792	0.0512
10	11.8722	0.0673	2.96	37.19	0.0795	0.0527

Table 4: Estimated value of a, b and λ at different values of T for women of ages at marriage 16 year and above from NFHS 4 data

T	Mean Conception	Homogeneous $\hat{\lambda}$	$\hat{a}$	$\hat{b}$	Heterogeneous $\hat{\lambda}$	$\hat{\lambda}_H$
5	9.9502	0.0557	2.02	37.88	0.0533	0.0269
6	10.6764	0.0591	2.36	38.78	0.0608	0.0350
7	11.2456	0.0610	2.37	35.67	0.0661	0.0381
8	11.7630	0.0616	2.54	36.69	0.0692	0.0419
9	12.1802	0.0620	2.46	34.05	0.0723	0.0428
10	12.6527	0.0614	2.78	38.50	0.0722	0.0462

A critical analysis of the above tables reveals that the estimate of homogeneous conception rate (λ) is almost lie in the vicinity of 0.067 for NFHS-3 and 0.061 for NFHS-4. There is not much variation with respect to T. The estimate of the harmonic mean of heterogeneous conception rate (λ_H) showing an increasing trend with respect to T. In NFHS-3 it is 0.0368 for $T = 5$ and increases to 0.527 for $T = 10$. For NFHS-4 estimate of λ_H is 0.0269 for $T = 5$ and increases to 0.0462 for $T = 10$. A similar trend is also observed in the estimate of heterogeneous λ. There is negligible variation in the estimate of λ under homogeneity assumption, but these estimates have variation under the assumption of heterogeneity. This clearly indicates that the conception rate is decreasing with respect to time.

5. Conclusion

Heterogeneity is sometimes used as a synonym for variability or diversity and places diculties in the way of interpretation of all statistical data based on averages. It introduces a clear bias in the mean as customarily estimated. That bias arises because of a selection effect. First birth interval has signied women's fertility at the early stage of married life and the start of parenthood. It is largely governed by fecundability (monthly chance of conception) i.e., unique features. There is less chance of recall lapse because it is the most important event in a woman's life. The proposed technique provides estimate adjusted for truncation and selection effect. The procedure is based on some simplifying assumptions. The incidence of fetal wastage, an important component of human reproduction, is not considered. The model may work better if the assumption of time variant conception rate is taken into consideration.

References

[1] Bhattacharya, B. N., Pandey, C. M., and Singh, K. K. (1988). Model for First Birth Interval and Some Social Factors. Mathematical Biosciences, 92(1):17-28.

[2] Chakraborty, K. C. (1976). Some Probability Distributions for Birth Intervals. PhD thesis, Banaras Hindu University, India.

[3] Gini, C. (1924). Premieres Recherches Sur La Fecondabilite De La Femme. In Proceedings of the International Mathematics Congress, Toronto.

[4] Henry, L. (1961). Fecondite et Familie-Models Mathematiques II Partie Theo-rique. Population, 16:27-48.

[5] Kumar, A. (2012). Sampling frame as a determinant of distribution of birth intervals. PhD thesis, Banaras Hindu University.

[6] Kumar, A. and Yadava, R. C. (2014). Analysisng censored data onfirst birth interval for heterogeneous group of females. J. of Combinatorics, Information System Sciences, 39(4):149-165.

[7] Kumar, A. and Yadava, R. C. (2015). Usual Closed Birth Interval versus Most Recent Closed Birth Interval. Journal of Data Science, 13:73-94.

[8] Kumar, A., Yadava, R. C., and Srivastava, U. (2010). Analysing the Impact of Marital Duration on First Birth Interval. Journal of Statistics and Application, 5(3-4):211-218.

[9] NFHS-3 (2007). National Family Health Survey-3, India. International Institute for Population Sciences, Mumbai, India.

[10] NFHS-4 (2016). National Family Health Survey-4, India. International Institute for Population Sciences, Mumbai, India.

[11] Potter, R. G. and Parker, M. (1964). Predicting the time required to concieve. Population Studies, 18:99-116.

[12] Pratap, M. (2011). A Study on Fertility Changes Through Birth Interval Ap-proach. PhD thesis, Banaras Hindu University.

[13] Sheps, M. C. (1964). On the Time Required for Conception. Population Studies, 18:85-97.

[14] Sheps, M. C. and Menken, J. A. (1972). Distribution of birth intervals according to the sampling frame. Theoretical Population Biology, 3(1):1-26.

[15] Sheps, M. C. and Menken, J. A. (1973). Mathematical models of conception and birth. The University of Chicago Press, Chicago and London.

[16] Sheps, M. C., Menken, J. A., Ridley, J. C., and Lingner, J. W. (1970). Truncation effect in closed and open birth interval data. Journal of the American Statistical Association, 65(330):678-693.

[17] Singh, S. N. (1964). A Probability Model for Couple Fertility. Sankhya: The Indian Journal of Statistics, Series B , 26((1/2)):89-94.

[18] Singh, S. N., Yadava, R. C., and Pandey, A. (1979). On a generalised probability distribution of open birth interval regardless of parity. Journal of Scientific Research, 29:162-170.

[19] Wolfers, D. (1968). Determinants of birth intervals and their means. Population Studies, 22(2):253-262.

[20] Yadava, R. C., Kumar, A., and Pratap, M. (2013). Estimation of Parity Pro-gression Ratios