Demography India
Special Issue (2017), pp: 04-15
ISSN: 0970-454X

A Probabilistic Study of Heterogeneity of Sterility in Uttar Pradesh

Brijesh P. Singh
Shweta Dixit
Sandeep Singh

Abstract

A Sterile female has no children by their deliberate choice or because of biologically infertility. Females who are childless by chance are not considered to be sterile. The object is to estimate the proportion of sterile females in Uttar Pradesh indirectly based on the NFHS- I & III data, by separating the two types of childless females into sterile and fertile one. Method of moment has been used for fitting the truncated and inflated negative binomial distribution to the data obtained from the females of age group 20-35 years whose age at marriage is below 20 years. Based on the fitted distribution, the proportion of sterile females is estimated at 2.4 and 3.2 percent for all females of NFHS-I & III respectively. This estimate is much lower than the corresponding estimate of sterile females in the USA, which was estimated at 11%. The difference between the two can be due to some socio-cultural factors influencing the deliberate choice of females to have no children. In the urban area proportion of sterile females are more than the rural females, also it is observed that in Uttar Pradesh the proportion of sterile females are increasing over the period.

Introduction

Desire for motherhood is very usual phenomenon and almost worldwide. Historically as well as traditionally, motherhood for women has been seen as natural. In most developing countries including India there has not been any notable change as far as the desire for motherhood is considered. Childlessness or sterility to have no children on the other hand affects both men and women of reproductive age in all parts of the world. Often the ill effects of childlessness are more severe for women than for men. The childless women are subjected to the additional risks of social discrimination in many forms such as restriction on their participation in social celebrations for example, allowing husband to remarry. These things happen irrespective of whether it childlessness is due to her being infertile or because the husband is infertile, as matter of fact, male infertility has rarely been considered a factor in childlessness. Despite this, the problem of childlessness has been largely overlooked in favor of research and promotion of family planning. Childlessness research has been neglected both as a health problem and as subject for social science research as in the past few decades greater amount of emphasis is placed on controlling the unwanted fertility.

The demographers and population scientists over the world have paid more concentration on trying to understand the dynamics of fertility and somehow ignored the important issue of childlessness to a greater extent and the Indian demographic community is also following the same trend. As a result very little work has been carried out in the past on this important aspect. Thus there is need to explore this rarely explored phenomenon. In some of the regions infertility is found to be widespread and its prevalence reaching such proportions that it can well be considered as a public health problem affecting the life of the whole society (WHO 1991). In its extreme, infertility, compounded by pregnancy wastage, infant and child mortality, may lead to depopulation, which poses serious threat to the social and economic development of the region. It is important to remind here that such high increase in the childlessness and also the fact that the levels of childlessness are higher in rural areas as compared to the urban areas, raises many concerns in view of the fact that the voluntary childless in India and its state is said to virtually non-existent.

Objective

Since childbearing is highly valued in terms of Indian context and childlessness can have devastating consequences for Indian women, infertility is perceived to be a very serious problem (Pachauri, 1998). So the distribution of completed number of children per female has become the subject of interest for all those who come in the category of demographers, human biologists, geneticists and social scientists. Most frequencies of such event occurrences i.e. count data can be described initially by Poisson distribution. But fail to capture the heterogeneity of the phenomenon as it varies from female to female. One of the crucial questions in statistical analysis of count data is how to formulate an adequate probability model to describe observed variation of counts. The very first choice for explaining these count phenomena was Poisson distribution or we can say the Poisson model is a benchmark model for the statistical analysis of these count data. Sometimes count data exhibit variation, referred to as over-dispersion or under-dispersion, resulting in the lack of fit of the Poisson model. In other words the major drawback of this distribution is the fact that the variance is restricted to be equal to the mean, a situation that may not be consistent with the observation here, i.e. the phenomena are characterized by two features: first is over-dispersion, i.e. the variance is greater than the mean; second is Zero-inflated, i.e. the percentage of zero values in the empirical distribution. This family made of distributions indexed by a positive parameter such the probability mass function is defined as

$$p_x = \frac{e^{-\lambda}\lambda^x}{x!}; x = 0,1,2,...,\infty; \lambda > 0 \tag{1}$$

Brass (1958) has also argued for the same while describing the probability of having *x* births. He state that if all women in a population had the same, constant expectation of bearing children per unit of time, for the same fixed period exposed to risk, the births would be spread randomly over the period and the distribution of women of completed fertility by number of births would follow the ordinary Poisson form. In brass's model constant λ signifies an individual's expectation of bearing children. But it is not the case in actual population as some women are not exposed to child bearing and the value of λ varies from individual to individual.

It follows that this family of distributions is a natural exponential family with canonical parameter $\theta = ln\ (\lambda)$ and cumulant function $k\ (\theta) = exp\ (\theta)$. One of the important features of the Poisson family is that the variance-to-mean ratio, also called Fisher dispersion index, is equal to 1 whatever the value of λ. Then, the Fisher dispersion index of a counts probability distribution is considered as a measure of its departure from Poisson model. Notice that the case where the variance-to-mean ratio equals to 1 characterizes the Poisson family among the natural exponential family of discrete distributions. Over-dispersion with respect to Poisson model refers to the cases where there is evidence that the observed random variation is greater than the expected random variation under the Poison model. Otherwise under-dispersion means that the expected variation is greater than the observed one. Another important feature of the Poisson family is the equality

$$z_i = \frac{1}{\lambda}\ln\left[p\left(0,\lambda\right)\right] + 1 = 0 \tag{2}$$

Where $p(0,\lambda)$ is the probability of zero. The index $z_i = \frac{1}{\lambda}\ln\left[p\left(0,\lambda\right)\right] + 1$, called zero inflation index, is also used as a measure of departure from Poisson model. The reliance of the Poisson model on a single parameter results in a lack of flexibility in its application. The lack of fit of the Poisson model is a frequent issue in the count data analysis literature as a survey can show. These results in proposals of alternative statistical analysis framework that take into account the knowledge on random mechanism underlying the occurrences of the counted events. But one has to notice that more attention has been paid to over-dispersion. The Negative binomial distribution is one of the most widely used distributions when modeling count data that exhibit variation that Poisson distribution cannot explain.

In view of this, alternative probability distribution, such as the negative binomial and generalized Poisson among others are preferred for modeling the phenomenon under study. The negative binomial distribution is a natural and more flexible extension of the Poisson distribution and allows for over-dispersion relative to the Poisson. The negative binomial distribution can be derived from several models. Accordingly, there are a variety of definitions in the literature. Though typically derived as a generalization of the geometric distribution, the negative binomial can also be derived as a mixture of Poisson distributions. Applications of the negative binomial distribution are wide-ranging. Dandekar (1955) developed modified Poisson distribution to explain the variation in the number of births of a homogeneous group of females during a given period (0, T) of length T. Brass (1958) modified the Dandekar's model assuming that conception rate varies among women according to a Pearson Type III distribution and named it as Negative Binomial distribution. Considering only births to fertile women, he applied the truncated version of the model to fit the data of some selected countries, relating to the distribution of mothers of completed family size according to number of children born to them. But, this distribution failed to describe the data for countries of low fertility. In some situation where the worsen under study consist of several distinct groups with respect to fecundability and the average risk of conceptions among groups differ markedly, these models may not provide an adequate fit to observed data, especially when the period of observation is long. The negative binomial distribution has been shown to have applicability in accident statistics, birth–death processes, market research, econometrics, biometrics, and ecology, among others.

However, there are also some good empirical evidences that the distribution is nearly that of a negative binomial (Kojima and Kelleher, 1962). In this study the Heterogeneity of Childless females suggests that, the number of childless females is much greater than the expected number of childless females when they fit the negative binomial distribution to the observed frequencies of completed family size. And this idea comes from the paper of Waller et al. (1973) in his paper entitled Heterogeneity of childless families. This led to infer that the childless females are a mixture of two types of females:

- The first is biologically fertile and could have children, but by chance, didn't. This type of females should be a part of general negative binomial distribution of family size and can be explained by this.
- The second is either biologically or electively not fertile or sterile and thus has no children. This should not be a part of the general negative binomial distribution of no. of children to a female.

The proportion of this type of females is expected to vary among populations studied, due to socio-cultural factors influencing the deliberate choice to have no children. Further, the purpose of this chapter is to test the adequacy of the general negative Binomial and its modified model in the data of Uttar-Pradesh and estimating the proportion of sterile females through different methodologies.

Probability Model

Model-I

For the present study Inflation Index (2.2) (Z_i =0.322034) measures a significant departure from Poisson model. In view of this, alternative probability distribution, such as the negative binomial has been applied and can be derived from several models. One way out of that is by the mixing of the Poisson with gamma distribution where the Poisson parameter λ describes the heterogeneity factor which Poisson distribution failed to express.

As the earlier discussion suggests that a birth process leads to a negative binomial distribution and it can be derived from the mixture of Poisson distribution. So, in order to derive it we suppose that a random variable $X \sim poisson(\lambda)$ and that λ itself is a random variable with $\lambda \sim gamma(\alpha,\beta)$. The argument is given by both Brass and Dandekar. They assumed that the population was composed of women with different expectations of bearing children, represented by different values of λ in a Poisson distribution. So, the unconditional distribution of X is negative binomial, as follows. Let $f(\lambda)$ denote the distribution of λ. Then

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, x = 0,1,2...$$

$$\& \ f(\lambda) = \left[\Gamma(\alpha)\beta^\alpha\right]^{-1}\left[\lambda^{\alpha-1}e^{-\lambda/\beta}\right], \lambda > 0 \qquad (3)$$

So, $p(x) = \int_0^\infty p(x|\lambda)\, f(\lambda)\, d\lambda$

$$= \left[x!\Gamma(\alpha)\beta^\alpha\right]^{-1}\int_0^\infty \lambda^{(x+\alpha)-1}e^{-\lambda(1+1/\beta)}d\lambda$$

$$= \frac{\Gamma(x+\alpha)\beta^x}{x!\Gamma(\alpha)(1+\beta)^{x+\alpha}}$$

$$= \binom{x+\alpha-1}{\alpha-1}\left(\frac{\beta}{1+\beta}\right)^x\left(\frac{1}{1+\beta}\right)^\alpha$$

Here, Let, $\frac{1}{1+\beta} = m$ and $\alpha = k$ for simplifying the above density.

Then, the negative binomial random variable *X* is a non-negative discrete random variable and the distribution is characterized by two parameter *m* and *k*, where is typically termed as the negative binomial dispersion parameter. The probability function for the negative binomial distribution with parameter *(m, k)* is given by

$$p(x|m,k) = \frac{(x+k-1)!}{(k-1)!(x)!} m^k (1-m)^x ; x = 0,1,2,... \tag{4}$$

$$where,\ m \in (0,1), n = (1-m), k > 0$$

The mean and variance of the negative binomial distribution are given below:

$$Mean = E(X) = \mu = \frac{k(1-m)}{m},$$

$$Variance = V(X) = \sigma^2 = \frac{k(1-m)}{m^2}$$

Note that here the variance is always greater than the mean i.e. a situation of under dispersion, i.e. $\mathrm{Var(X)} > E(X)$. The negative Binomial distribution, in the present context, is based on the following assumptions:

1. All the women are exposed to risk of childbearing for the same period *T*.
2. Each woman has an expected rate of childbearing *X*, which is constant over *T*.
3. The distribution of *X* among the women is given by equation [3]
4. The births occur at random.

The constraint arising from the first assumption is that the model be tested in a narrow age range, preferably, close to the end of childbearing; hence our choice of the age group 45-49. Assumptions 2 and 4 are more likely and acceptable. It would appear that the assumed distribution of *X* in the population should receive a harder look.

Model-II

The second probability model for the representations of extensive data sets of children per females is given as follows.

1. At any point in time, let *α* be the proportion of fertile females and *(1-α)* be the proportion of sterile females.
2. If *k* represents the risk of birth and the pattern of births follows the modified negative binomial distribution.

If *x* represents number of births or child per female has a modified negative binomial distribution with probability function,

$$p(x|m,k) = \begin{cases} P(X=0) = 1-\alpha+\alpha\ m^k & ;x=0 \\ P(X=\mathrm{x}) = \alpha\ \dfrac{(x+k-1)!}{(k-1)!(x)!} m^k (1-m)^x & ;x=1, 2, 3,.... \end{cases} \tag{5}$$

$$where,\ m \in (0,1), n = (1-m), k > 0 \,\&\, 0 < \alpha < 1$$

The above distribution (5) is termed as Inflated negative binomial distribution. The mean and variance of the distribution are given below:

$$Mean = E(X) = \mu = \frac{\alpha k(1-m)}{m}$$

$$Variance = V(X) = \sigma^2 = \frac{\alpha k(1-m)}{m^2}\left[1+(1-\alpha)k(1-m)\right]$$

The first term of the distribution is of particular interest to the problem under consideration is the childless fertile families, $x = 0$ and $Pr[x = 0] = p_0 = p^k$, which is the theoretical proportion of childless fertile couples. Hence to estimate this proportion we need to obtain estimates for *m* and *k*.

This Model-I is found to be a useful distribution as it has its representation in several fields of demography. Furry (1937) and Kendall (1949) have shown its application in birth and death process. In numbers of modeling for count is given by Waller et al. (1973), Wilson *et al.* (1983), Kault (1996) and Binns (1986). With this in mind, two methods of fitting this distribution, to the number of children per female data obtained from the National Family Health Survey-III (2005-2006), are employed. The best fitted distribution is used to estimate (or approximate) the proportion of sterile couples in our population. As the issue is very sensitive, there has not been any attempt to estimate this quantity directly; to the best of our knowledge.

Source of Data

Data for this study is taken from the third round of National Family Health Survey (NFHS-III), which was carried out in 2005-2006 from the IIPS, Mumbai. We are working with the Uttar-Pradesh population. The observed distribution of females for number of children is shown in table 1. The sample size for the present analysis is 2221 and 4180 married women of Uttar Pradesh whose marital age was less than 20 years and their current age is between 20-35 years for NFHS-III and NFHS-I respectively.

Estimation Methods

Here are the methods of fitting the negative binomial distribution to the observed data have been employed (see Waller *et al.*, 1973).

- Method of Moments (Model-I)

This method consists of approximating the mean (μ) and the variance (σ^2) of the negative binomial distribution directly from the observed data, and the parameters *m*, *k* are estimated using the formulas:

$$m = \frac{\mu}{\sigma^2}, \; k = \frac{(\mu m)}{(1-m^2)}$$

- Rider Method of Estimation (Model-I)

In this second method we use the same method to develop the formulas for estimating the parameter of a truncated negative binomial distribution. In this method the zero class is considered as missing and the parameters *m*, *k* is estimated on the basis of the incomplete (truncated) distribution. The formula came from Rider (1955). Let

$$T_i = \sum_{x=0}^{\infty} x^i f_x$$

Where f_x is the frequency of families of size *x*. All summations range from $x = 0$ to the maximum class value which is to be possible (i.e. number of children to a female). Thus,

$$\left.\begin{aligned} T_0 &= \sum_{x=0}^{\infty} f_x \\ T_1 &= \sum_{x=0}^{\infty} x^1 f_x \\ T_2 &= \sum_{x=0}^{\infty} x^2 f_x \end{aligned}\right\} \tag{6}$$

Further, let

$$\left.\begin{aligned} T_0' &= N\sum_{x=0}^{k-1} f_x + T_0 \\ T_1' &= N\sum_{x=0}^{k-1} x f_x + T_1 \\ T_2' &= N\sum_{x=0}^{k-1} x^2 f_x + T_2 \end{aligned}\right\} \tag{7}$$

We know that the general term in the binomial expansion $(r-k)^{-x}$, in which $r=1+k$. As we wish to estimate *m* and *k*, also the number *N* in a sample before truncation, we need to use three moments. The first three moments about origin are;

$$\mu_1' = km$$

$$\mu_2' = km(1+m+km)$$

$$\mu_3' = km(1+3m+2m^2+3km+3km^2+k^2m^2) \quad (8)$$

We shall consider only the case in which the class corresponding to $x=0$ has been truncated for the general case are complicated to be of interest. In the notation of formulas (6) to (7) with $k=1$, we have, since, $p_0=(1+p)^{-k}$

$$T_0' = N(1+p)^{-k} + T_0; T_i' = T_i,\ i>0 \quad (9)$$

Now T_2'/T_1' is an estimate of μ_2'/μ_1' and consequently we set

$$T_2 = T_1(1+p+kp) \quad (10)$$

$$T_3 = T_1(1+3p+2p^2+3kp+3kp^2+k^2p^2) \quad (11)$$

Next we solve (9) for *km* and substitute in (10).

Solving the resulting equation, we find the estimates of the parameters for the negative binomial distribution;

$$\hat{k} = \frac{(2T_2^2 - T_1T_2 - T_1T_3)}{(T_1T_3 - T_1T_2 + T_1^2 - T_2^2)} \quad (12)$$

$$\hat{m} = \frac{(T_1T_2 - T_1^2)}{(T_1T_3 - T_2^2)} \quad (13)$$

An estimate of *N* can now be obtained from the equation

$$N - T_0 = Nr^{-m} \quad (14)$$

- Method of Moments (Model-II)

Inflated negative binomial distribution has two parameters α and *k* to be estimated. As the mean and variance of the given density is as follows,

$$Mean = E(X) = \mu = \frac{\alpha k(1-m)}{m}$$

(15)

$$Second\ moment = E(x^2) = \frac{\alpha k(1-m)}{m^2}\left[1+k-mk\right]$$

(16)

$$Variance = E(x^2) - [E(x)]^2 = \frac{\alpha k(1-m)}{m^2}\left[1+(1-\alpha)k(1-m)\right]$$

(17)

And the zero cell frequency,

$$p(0) = 1-\alpha+\alpha m^k = \frac{n_0}{N}$$

(18)

Now we have

$$\mu_2' = \frac{\mu_1'}{m}\left[1+k-mk\right] \Rightarrow \frac{m\mu_2'}{\mu_1'} = \left[1+k-mk\right] \Rightarrow k = \frac{m\mu_2' - \mu_1'}{\mu_1'(1-m)} \quad (19)$$

and

$$\Rightarrow \alpha = \frac{m\mu_1^{'}}{k(1-m)} \tag{20}$$

Also we have,

$$\frac{n_0}{N} = 1 - \alpha + \alpha m^k$$

Putting the values of *α* and *k* we get,

$$\frac{n_0}{N} = 1 - \left(\frac{m\mu_1^{'}\left(\mu_1^{'}(1-m)\right)}{\left(m\mu_2^{'} - \mu_1^{'}\right)(1-m)} \right) + \left(\frac{m\mu_1^{'}\left(\mu_1^{'}(1-m)\right)}{\left(m\mu_2^{'} - \mu_1^{'}\right)(1-m)} \right) m^{\left(\frac{m\mu_2^{'}-\mu_1^{'}}{\mu_1^{'}(1-m)}\right)}$$

$$\Rightarrow 1 + \frac{n_0}{N} - \left(\frac{m\mu_1^{'2}}{\left(m\mu_2^{'} - \mu_1^{'}\right)} \right) + \left(\frac{m\mu_1^{'2}}{\left(m\mu_2^{'} - \mu_1^{'}\right)} \right) m^{\left(\frac{m\mu_2^{'}-\mu_1^{'}}{\mu_1^{'}(1-m)}\right)} \tag{21}$$

Solving the above equation iteratively gives us the corresponding values of the estimates *α* and *k*.

Analysis & Findings

- Method-I

As Table 1 represents the observed number of children per female for our urban, rural and total population of Uttar Pradesh separately (U, R, T respectively) from NFHS-III and NFHS-I. In table1, for all females with number of children 11 and more, we are considering it 11+ and the table is as follows:

Table 1: Distribution of no. of Children per female in Uttar Pradesh for NFHS-III & NFHS-I

No. of children per female(x)	Urban population		Rural population		Total population	
	NFHS-III	NFHS-I	NFHS-III	NFHS-I	NFHS-III	NFHS-I
0	62	62	160	347	222	409
1	111	105	277	630	388	735
2	99	127	353	748	452	875
3	78	87	320	660	398	747
4	52	60	241	490	293	550
5	53	53	183	339	236	392
6	25	37	107	212	132	249
7	9	15	56	124	65	139
8	3	7	18	49	21	56
9	2	3	7	13	8	15
10	1	2	2	9	4	12
11	0	0	2	1	2	1
Total	**495**	**558**	**1726**	**3622**	**2221**	**4180**

From Table 1 we can conclude that contains the childless females is approximately 10 percent (9.99%) for Uttar Pradesh. Separately this proportion in urban and rural setting is 12 and 9.2 percent from the NFHS-III data. Furthermore, the total number of electively or biologically sterile females is approximated at 222 which is the actual no. of childless female rather than sterile females and should be noted that not all childless females (crude estimate) would give a number that is higher than the actual number. So, further for approximating that proportion of sterile females, we approximate the first cell as missing in the next method proposed by Rider (1955), as we consider the childless females being a mixture of two types of females, one biologically fertile and could have children but did not this type of type of females should be a part of the general negative binomial distribution and two, biologically or electively not fertile and thus has no children.

- Method-II

Here we deal with the zero class as missing values and for this we use Rider Method (1955) to estimate the parameters on the basis of the incomplete distribution. Let just have the Uttar Pradesh data for example, hence

$$T_i = \sum_{x=1}^{11} x^i f_x, i = 0,1,2,3$$

Then we have, T_0 = 1999, T_1= 6387, T_2 = 26937, T_3=137055
Then,

$$\hat{k} = \frac{(2T_2^2 - T_1T_2 - T_1T_3)}{(T_1T_3 - T_1T_2 + T_1^2 - T_2^2)} = 21.80809 \qquad \hat{m} = \frac{(T_1T_2 - T_1^2)}{(T_1T_3 - T_2^2)} = 0.876373$$

$$\hat{m}^{\hat{k}} = 0.056254 \qquad \tilde{N} = \frac{T_0}{1-\hat{m}^k} = 2118.154 \qquad \hat{N}_0 = \hat{m}^{\hat{k}}\tilde{N} = 119.1542$$

Using these values for the population the fitted distribution will be as follows:

$$f(x) = \frac{(x+21.80809-1)!}{(21.80809-1)!\,(x)!}(0.876373)^{21.80809}(1-0.876373)^x \;,\; for \forall x$$

And so on for the distributions of rural and total population can also be done and the following table can be prepared.

Table 2 contains the No. of children per female distribution based on Model-I. In the table column 1 contains the No. of children per female (*x*). Column 2, 4 and 6 contains the observed number females for each value of *x*, in the population of urban, rural and total Uttar Pradesh with the zero class being adjusted. In the table column 1 contains the No. of children per female (*x*). Column 2, 4 and 6 contains the observed number females for each value of *x* and Column 3, 5 and 7 contains the expected no. of females of each value of *x* based on the negative binomial distribution of U, R, T population respectively. It can be seen from this table 2 that the fitted data and the actual data have a very similar general shape with some discrepancies in the empirical and the theoretical probabilities. The total number of electively or biologically sterile females is approximated at 2221−2117.15=103.85≅104. In other words, based on this method about 46.77% of all childless females are electively or biologically sterile.

Table 2: Distribution of no. of Children per female in Uttar Pradesh by incomplete (Rider) method for NFHS-III (Model-I)

No. of children per female (x) [1]	Urban population		Rural population		Total population	
	Observed number of females f(x) [2]	Expected number of females [3]	Observed number of females f(x) [4]	Expected number of females [5]	Observed number of females f(x) [6]	Expected number of females [7]
0	35.42	42.66	84.49	94.76	119.15	135.96
1	111	90.15	277	250.74	388	341.02
2	99	105.49	353	349.85	452	456.38
3	78	90.29	320	342.28	398	432.75
4	52	63.09	241	263.52	293	325.95
5	53	38.14	183	169.92	236	207.36
6	25	20.65	107	95.40	132	115.74
7	9	10.26	56	47.88	65	58.16
8	3	4.75	18	21.89	21	26.79
9	2	2.08	7	9.25	8	11.47
10	1	0.86	2	3.65	4	4.61
11	0	0	2	1.36	2	1.76
Total	468.4237	468.4237	1650.497	1650.497	2118.154	2117.154
$\hat{m}$	0.77269		0.85551		0.83187	

$\hat{k}$	9.2960		18.3134		14.9193	
K-S Statistics	D_{cal} $D_{0.05}$	0.015547 0.040122	D_{cal} $D_{0.05}$	0.01844 0.03571	D_{cal} $D_{0.05}$	0.019862 0.035417

This method is giving proportion of sterile but the logical reasoning behind the methodology is not clear. So for clear representation of the proportion of sterile here we, propose an Inflated Negative binomial distribution incorporating another parameter for judging the sterility proportion more accurately. Further, the third method which is explained above is giving the proportion sterile in the following table 3 for the NFHS-III data of Uttar Pradesh. It can be seen from this table 3 that the fitted data and the actual data have a very similar general shape than the previous model. The total number of electively or biologically sterile females is approximated at 1-$\hat{\alpha}$ = 0.032. In other words, based on this method through inflated model about 3.2% of all females are electively or biologically sterile. This is very less than the previous estimates of sterile proportion of females. The suitability of the proposed distribution is verified by K-S statistics.

Table 3: Distribution of no. of Children per female in Uttar Pradesh by Inflated model for NFHS-III (Model-II)

No. of children per female (x) [1]	**Urban population**		**Rural population**		**Total population**	
	Observed number of females f(x) [2]	**Expected number of females [3]**	**Observed number of females f(x) [4]**	**Expected number of females [5]**	**Observed number of females f(x) [6]**	**Expected number of females [7]**
0	62	62.00	160	160.00	222	222.00
1	111	118.08	277	263.59	388	362.16
2	99	115.55	353	356.37	452	466.10
3	78	86.16	320	341.28	398	430.68
4	52	54.20	241	259.54	293	319.78
5	53	30.31	183	166.68	236	202.61
6	25	15.54	107	93.90	132	113.66
7	9	7.45	56	47.61	65	57.87
8	3	3.38	18	22.13	21	27.21
9	2	1.47	7	9.56	8	11.97
10	1	0.62	2	3.88	4	4.98
11	0	0.25	2	1.49	2	1.97
Total	495	495.00	1726	1726.00	2221	2221.00
$\hat{m}$	0.7204		0.831		0.802	
$\hat{k}$	6.8040		15.0393		12.0324	
$\hat{\alpha}$	0.9799		0.9670		0.9681	
K-S Statistics	D_{cal} $D_{0.05}$	0.06868 0.07326	D_{cal} $D_{0.05}$	0.01725 0.03923	D_{cal} $D_{0.05}$	0.03191 0.03459

Table 4: Distribution of no. of Children per female in Uttar Pradesh by incomplete (Rider) method for NFHS-I (Model-I)

No. of children per female (x) [1]	**Urban population**		**Rural population**		**Total population**	
	Observed number of females f(x) [2]	**Expected number of females [3]**	**Observed number of females f(x) [4]**	**Expected number of females [5]**	**Observed number of females f(x) [6]**	**Expected number of females [7]**
0	38.14	44.41	204.98	232.68	243.82	277.24
1	105	95.18	630	566.17	735	662.20
2	127	114.55	748	743.85	875	858.84
3	87	101.98	660	699.72	747	801.34
4	60	74.82	490	527.67	550	601.90
5	53	47.86	339	338.85	392	386.39

6	37	27.62	212	192.31	249	219.92
7	15	14.70	124	98.89	139	113.74
8	7	7.33	49	46.90	56	54.39
9	3	3.46	13	20.78	15	24.36
10	2	1.57	9	8.69	12	10.32
11	0	0.68	1	3.46	1	4.17
Total	534.1458	534.1458	3479.985	3479.985	4014.819	4014.819
$\hat{m}$	0.7362		0.8056		0.7947	
$\hat{k}$	8.1258		12.5167		11.6349	
K-S Statistics	D_{cal} $D_{0.05}$	0.029982 0.070527	D_{cal} $D_{0.05}$	0.011578 0.027631	D_{cal} $D_{0.05}$	0.013834 0.025725

Table 5: Distribution of no. of Children per female in Uttar Pradesh by Inflated model for NFHS-I (Model-II)

No. of children per female (x) [1]	Urban population		Rural population		Total population	
	Observed number of females f(x) [2]	Expected number of females [3]	Observed number of females f(x) [4]	Expected number of females [5]	Observed number of females f(x) [6]	Expected number of females [7]
0	62	62.00	347	347.00	409	409.00
1	105	118.75	630	637.24	735	786.68
2	127	124.90	748	781.58	875	924.34
3	87	100.08	660	697.17	747	796.48
4	60	67.67	490	505.27	550	561.52
5	53	40.67	339	315.49	392	343.09
6	37	22.41	212	175.89	249	188.13
7	15	11.54	124	89.65	139	94.73
8	7	5.64	49	42.48	56	44.53
9	3	2.63	13	18.95	15	19.76
10	2	1.19	9	8.03	12	8.36
11	0	0.52	1	3.26	1	3.39
Total	558	558.00	3622	3622.00	4180	4180.00
$\hat{m}$	0.6995		0.777		0.766	
$\hat{k}$	6.6106		10.3648		9.6845	
$\hat{\alpha}$	0.9812		0.9755		0.9758	
K-S Statistics	D_{cal} $D_{0.05}$	0.062606 0.069003	D_{cal} $D_{0.05}$	0.027556 0.027084	D_{cal} $D_{0.05}$	0.045033 0.025212

Table 6: Estimated proportion of sterile females through fitted Negative Binomial/Inflated Negative Binomial Distribution for NFHS-III & I

Place of residence	NFHS-III			NFHS-I		
	Crude estimate	Model-I	Model-II	Crude estimate	Model-I	Model-II
Urban	0.13	0.057	0.020	0.11	0.040	0.019
Rural	0.09	0.046	0.033	0.10	0.041	0.025
Total	**0.10**	**0.049**	**0.032**	**0.10**	**0.041**	**0.024**

Results & Conclusions

The problem of trying to construct models to describe the distribution of births in human populations has engaged the attention of demographers and statisticians for a long time. Using data on the distribution of number of children per women obtained from NFHS in Uttar Pradesh we tried to find suitable probability distributions to fit the observed frequency distributions. The fitted ordinary Negative Binomial distributions (by Rider method), as well as modifications of this as Inflated Negative Binomial distributions have been presented in this chapter. Here the truncated negative

Binomial (truncated at zero) and Inflated negative Binomial to observed distributions of number of children per women. The fit was satisfactory in some cases and unsatisfactory in others, but the exercise enabled us to identify patterns of discrepancy between observed and fitted distributions and also to get the "proportion of sterile women" in the population under study. Thus the number of number of children per women follows, approximately, the "Negative Binomial form" as well as its modified form. The observed and fitted distributions of women by number of children per women are made in above table 1 and table 3 for Uttar Pradesh through negative binomial, truncated and inflated negative binomial respectively for NFHS-III. For comparison point of view data of NFHS-I has also been taken and shown in table 4 and 6 respectively for the two models. The fitting was performed separately for Uttar Pradesh (Urban, Rural and total) in the same tables. Apart from showing the observed and fitted number of women responding to number of children per women, the tables also show the values of the two parameters of the negative Binomial distribution as well as the computed K-S statistics for the comparison.

On the other hand the method of fitting the negative binomial distribution to the population data gives us the idea how to estimate the proportion of two types of childless i.e. sterile and fertile females in population of Uttar Pradesh. It should be considered that all childless females should not be considered as sterile females. Thus estimating the proportion of sterile females based on all childless females gives us a crude estimate and must be higher than the actual number. So for approximating that proportion, we consider the childless females as being a mixture of biologically fertile and could have children but did not, this type of females should be a part of the general negative binomial distribution of family size and biologically or electively not fertile and thus has no children.

Since model II is theoretically better and giving us the proportion of sterile females by incorporating the parameter of sterility than I, we may conclude that 0.032 is a better and reasonable estimate of proportion sterile. Hence, the percentage of electively or biologically sterile couples in Uttar-Pradesh population is about 3.2 percent as a whole from NFHS-III and the proportion is increased for Uttar Pradesh population as the proportion from NFHS-I is 2.4 percent. So, we can say in the last decade the proportion of sterility is increased about 0.8 percent. This is an obvious result as we can verify this by the increasing number of infertility patents in the present time.

References

Binns, M. R. (1986). Behavioral Dynamics and the Negative Binomial Distribution. *Oikos*, 47, 315-318.20, 133-138.

Brass, W. (1958). The distribution of births in human population, *Population Studies*, 12, 51-72.

Dandekar, V. M. (1955). Certain modified forms of binomial and poisson distributions, Shankhya, Series B 12 237-250.

Furry, W.H. (1937). On Fluctuation Phenomena in the Passage of High Energy Electron Through leads. *Physical Review*, **52**, 569-581.

Kault, D. (1996). The Shape of the Distribution of the Number of Sexual Partners. *Statistics in Medicine*, **15**, 221-230.

Kendaal, M.G. (1949). Stochastic Process and Population Growth. *Journal of Royal Statistical Society, Series B*, **11**, 230-282.

Kojima, K.I. and Kelleher,T.M. (1962) Survival of Mutant Genes. *Amer. Nature*, 329-326.

Md Fraiwan Al-Saleh and F. K. AL-Batainah (2003). Estimation of the Proportion of Sterile Couples Using the Negative Binomial Distribution, Journal of Data Science 1,(2003), 261-274.

National Family Health Survey-I (1992-93). International Institute of Population Sciences, Mumbai.

National Family Health Survey-III (2005-06). International Institute of Population Sciences, Mumbai.

Waller, T.H, Rao, B.R, and Li, C.C. (1973). Heterogeneity of Childless Families. *Social Biology*,

Rao, B.R., Mazumdar, S., Waller, T.H, and Li, C.C. (11973). Correlation Between the Number of Two Types of Children in a Family. *Biometrics*, **29**, 271-279.

Rider, P.R. (1955). Truncated Binomial and Negative Binomial Distribution. Journal of the American Statistical Association, 50, 877-883.

WHO Technical Report Series (No. 442). Health Aspects of Family Planning, WHO Geneva.

Wilson, L.J., Fulks, J.L., and Young, J.H. (1984). Multistage estimation compared with fixed sample Size Estimation of The negative Binomial Parameter k. *Biometrics*, **40**, 109-117.

Dr. Brijesh P. Singh
Associate Professor
Department of Statistics
Institute of Science
Banaras Hindu University, Varanasi, India

Dr. Shweta Dixit
Research Assistant Professor
School of Management
SRM University, Kattankulathur
Tamil Nadu, India

Sandeep Singh
Research Scholar
Department of Statistics
Institute of Science
Banaras Hindu University, Varanasi, India

Corresponding Author: Dr. Shweta Dixit
Email: shwetadixit.s@ktr.srmuniv.ac.in