Chapter 2

METHODOLOGY, DATA AND ANALYTICAL FRAMEWORK

In the absence of reliable civil registration data for states as well as for the country as a whole, demographers largely rely on census or survey data to provide the estimates of fertility and mortality at various levels in order to understand population dynamics and the impact of on-going health and family welfare programme. The importance of survey data is all the more with a realization that in a dynamic population a census at infrequent intervals is of limited use. Even highly precise information about the component of a population from various consecutive censuses that might be at interval of say 10 years, may not help much in planning that demands a knowledge of the population for the very recent period. A series of surveys with small sample at a reasonably shorter interval may be more useful. As confidence in sampling has increased, the practice of relying on samples for the collection of important series of fertility and mortality data that are published at regular intervals has become common. For example, Sample Registration System (SRS), maintained by the Office of the Registrar General of India, which is a large scale demographic sample survey based on the mechanism of dual record system, is in operation in the country since more than three decades to provide reliable estimates of fertility and mortality at the state and national level for rural and urban areas separately at regular intervals of one year. Similarly, special surveys including National Sample Survey (NSS) and National Family Health Survey (NFHS) are carried out in the country from time to time with the aim to provide highly precise information about the marriage pattern, fertility, mortality, health and other aspects of the population. However, such surveys with moderate sample size may not be able to provide reliable demographic parameters, particularly estimates of mortality by age and sex at the state or lower levels because of large sampling error involved in such estimation, apart from the error of incompleteness of reporting of births and deaths in retrospective surveys.

When the same population is sampled repeatedly (apart from the changes that the passage of time introduces), the sampler is in an ideal position to make realistic estimates both of costs and of variances and to apply the techniques that lead to optimum efficiency of sampling (Cochran, 1966). However, many considerations affect the estimate. The question arises how frequently such surveys should be repeated and in what manner the sample should be changed as time progresses. Respondents may be unwilling to give the same type of information time after time. On the other hand, sometimes co-operation may be better in a second interview than in the first, and when the information is technical or confidential the second visit may produce more accurate data than the first. Nevertheless, our question is whether one can obtain a better estimate from the series of repeated samples. In most surveys, interest centres on the current estimate of a parameter, particularly when the characteristics of the population are likely to change rapidly with time. In such a situation, equal precision is obtained either by keeping the same sample or by changing it on every occasion, although replacement of part of the sample on each occasion may be better than these alternatives. On the other hand, it is best to retain the same sample throughout all occasions for estimating change in population parameters from one occasion to the next (Cochran, 1966, pp. 341-347). With a population in which time changes are slow, an average taken over all occasions may be adequate in most cases. However, for estimating the average over all occasions it is best to draw a new sample on each occasion. For example, this may be the case in a study of prevalence of chronic diseases of long duration. With a disease whose prevalence shows marked seasonal variation, the current data are of major interest, but annual averages are also useful for comparisons between different regions and over the years. Similarly, in a population in which certain other characteristics of the population are less likely to change rapidly with time, an average taken over all occasions during the last few years may provide a better estimate than that obtained from one occasion, as this may also reduce sampling error. Therefore, such an average taken over samples under this situation may be adequate. This aspect is investigated in the following section in light of availability of yearly SRS data and special national/state level surveys over time in India.

Merits of Repeated Sampling Technique:
A Theoretical Investigation

Suppose we were free to alter or retain the composition of the sample and the total size of sample may or may not be the same on all occasions. Therefore, there could possibly be three situations depending on the study design, in the repeated sampling of the same population; viz.

1) To retain the same sample throughout all occasions

2) To draw an independent (new) sample on each occasion (that is, changing it on every occasion)

3) To replace part of the sample on each occasion or there could be a situation where a portion of sample is matched (retained) and the remaining portion is unmatched (replaced by a new selection) on each occasion.

Let us examine each of these situations in succeeding sections.

Situation 1: When the same sample is retained on all occasions

Let $\overline{Y}_1$ and $\overline{Y}_2$ be the mean for a characteristic of the population, based on sample n_1 and n_2 on the first and second occasion respectively. Let the mean ($\overline{Y}_1$) of the first sample and mean ($\overline{Y}_2$) of the second sample have variances:

$$V(\overline{Y}_1) = S_1^2 = \frac{\sigma^2}{n_1} \text{ and } V(\overline{Y}_2) = S_2^2 = \frac{\sigma^2}{n_2}$$

Since the same sample is retained on both the occasions under this sampling plan,

$$n_1 = n_2 = n$$

Therefore, the over-all mean ($\overline{Y}$) for the two occasions is given by

$$\overline{Y} = \frac{(\overline{Y}_1 + \overline{Y}_2)}{2}.$$

In estimating the over-all mean ($\overline{Y}$) for the two occasions, the variance of $\overline{Y}$ is given by

$$V(\overline{Y}) = \frac{1}{4}V(\overline{Y}_1) + \frac{1}{4}V(\overline{Y}_2) + 2\frac{1}{4}\rho\sqrt{V(\overline{Y}_1)V(\overline{Y}_2)}$$

$$= \frac{1}{4}S_1^2 + \frac{1}{4}S_2^2 + \frac{1}{2}\rho S_1 S_2$$

$$= \frac{1}{4}(2S^2) + \frac{1}{2}\rho S^2, \text{ where } S_1^2 = S_2^2 = S^2 = \frac{\sigma^2}{n} \text{ and } n_1 = n_2 = n$$

$$= \frac{S^2}{2}(1+\rho)$$

$$\leq S^2 = \frac{\sigma^2}{n}, \text{ as } \rho \leq 1$$

Therefore, if the same sample is retained on both the occasions, the variance of the over-all mean for two occasions is expected to be less than that of either of the sample mean.

Situation 2: When an independent (new) sample is drawn on each occasion and the total size of the sample need not be the same on all occasions

Therefore, if samples are independent and have sizes n_1 and n_2, the over-all mean ($\overline{Y}$) for the two occasions is given by

$$\overline{Y} = \alpha\overline{Y}_1 + (1-\alpha)\overline{Y}_2, \text{ where } \alpha = \frac{n_1}{n_1+n_2}$$

$$V(\overline{Y}) = \alpha^2 S_1 + (1-\alpha)^2 S_2^2, \text{ where } S_1^2 = \frac{\sigma^2}{n_1},\ S_2^2 = \frac{\sigma^2}{n_2}$$

$$= \alpha^2 \frac{\sigma^2}{n_1} + (1-\alpha)^2 \frac{\sigma^2}{n_2}$$

$$= \frac{n_1^2}{(n_1+n_2)^2}\frac{\sigma^2}{n_1} + \frac{n_2^2}{(n_1+n_2)^2}\frac{\sigma^2}{n_2}, \text{ after substituting the value of } \alpha$$

$$= \frac{n_1\sigma^2 + n_2\sigma^2}{(n_1+n_2)^2}$$

$$= \frac{\sigma^2}{(n_1+n_2)} < \frac{\sigma^2}{n_i}, \text{ where } i = 1, 2$$

Therefore, even if independent sample of varied size is drawn on both the occasions, the variance of the over-all mean for the two occasions is expected to be less than that of either of the sample mean.

***Situation 3*:** When a part of the sample is matched (retained) in drawing sample from the same population on each occasion

Let $\overline{Y}_1$ be the mean of the first sample (of size n_1) and $\overline{Y}_2$ be the mean of the second sample (of size n_2). In selecting the second sample, n_3 are the units of the first sample which are matched (retained) (m stands for matched cases). The remaining units are unmatched portion of the sample (u stands for unmatched cases).

Let $\overline{Y}_{hm}$ be the mean of matched portion on occasion h (h = 1,2).

Let $\overline{Y}_{hu}$ be the mean of unmatched portion on occasion h (h = 1,2).

Let $\overline{Y}_h$ be the mean of whole sample on occasion h (h= 1,2).

Therefore,

$$\overline{Y}_1 = \frac{n_3\overline{Y}_{1m} + (n_1 - n_3)\overline{Y}_{1u}}{n_1}, \quad \overline{Y}_2 = \frac{n_3\overline{Y}_{2m} + (n_2 - n_3)\overline{Y}_{2u}}{n_2}$$

and $\overline{Y} = \frac{n_1\overline{Y}_1 + n_2\overline{Y}_2}{n_1 + n_2}$, where n_1 and n_2 are the total size of the first and second sample, and n_3 is the matched portion of the sample, while $n_1 - n_3$ and $n_2 - n_3$ are the unmatched portion of the first and second sample respectively.

In other words,

$\overline{Y}_1 = \alpha_1\overline{Y}_{1m} + (1-\alpha_1)\overline{Y}_{1u}$, $\overline{Y}_2 = \alpha_2\overline{Y}_{2m} + (1-\alpha_2)\overline{Y}_{2u}$,

where, $\alpha_1 = \frac{n_3}{n_1}$ and $\alpha_2 = \frac{n_3}{n_2}$

and $\overline{Y} = \alpha\overline{Y}_1 + (1-\alpha)\overline{Y}_2$, where $\alpha = \frac{n_1}{n_1 + n_2}$

$$= \alpha[\alpha_1\overline{Y}_{1m} + (1-\alpha_1)\overline{Y}_{1u}] + (1-\alpha)[\alpha_2\overline{Y}_{2m} + (1-\alpha_2)\overline{Y}_{2u}]$$

$$= \alpha\alpha_1\overline{Y}_{1m} + \alpha(1-\alpha_1)\overline{Y}_{1u} + (1-\alpha)\alpha_2\overline{Y}_{2m} + (1-\alpha)(1-\alpha_2)\overline{Y}_{2u}$$

Therefore,

$$V(\overline{Y}) = (\alpha\alpha_1)^2\frac{\sigma^2}{n_3} + \{\alpha(1-\alpha_1)\}^2\frac{\sigma^2}{n_1 - n_3} + \{\alpha_2(1-\alpha)\}^2\frac{\sigma^2}{n_3} + (1-\alpha)^2(1-\alpha_2)^2\frac{\sigma^2}{n_2 - n_3}$$

$$+2\alpha(1-\alpha)\alpha_1\alpha_2\rho\sqrt{\frac{\sigma^2}{n_3}\frac{\sigma^2}{n_3}}$$

$$= \sigma^2\left[\frac{\alpha^2\alpha_1^2}{n_3} + \frac{\alpha_2^2(1-\alpha)^2}{n_3} + \frac{2\alpha(1-\alpha)\alpha_1\alpha_2\rho}{n_3} + \frac{\alpha^2(1-\alpha_1)^2}{n_1 - n_3} + \frac{(1-\alpha)^2(1-\alpha_2)^2}{n_2 - n_3}\right]$$

$$\le \sigma^2\left[\frac{\{\alpha\alpha_1 + \alpha_2(1-\alpha)\}^2}{n_3} + \frac{\alpha^2(1-\alpha_1)^2}{n_1 - n_3} + \frac{(1-\alpha)^2(1-\alpha_2)^2}{n_2 - n_3}\right], \text{ as } \rho \le 1$$

$$= \sigma^2\left[n_3\left(\frac{2}{n_1 + n_2}\right)^2 + \frac{n_1 - n_3}{(n_1 + n_2)^2} + \frac{n_2 - n_3}{(n_1 + n_2)^2}\right],$$ when values of α, α_1 and α_2 are substituted

$$= \frac{\sigma^2}{(n_1+n_2)^2}\left[4n_3+n_1-n_3+n_2-n_3\right]$$

$$= \sigma^2 \frac{(n_1+n_2+2n_3)}{(n_1+n_2)^2}$$

Case 1: Let $n_1 = n_2 > n_3$

$$V(\bar{Y}) = \sigma^2 \frac{(n_1+n_2+2n_3)}{(n_1+n_2)^2}$$

$$= \sigma^2 \frac{(2n_1+2n_3)}{(2n_1)^2}$$

$$= \sigma^2 \frac{(n_1+n_3)}{2n_1^2}$$

$$= \frac{\sigma^2}{n_1}\left[\frac{n_1+n_3}{2n_1}\right]$$

$$< \frac{\sigma^2}{n_1} = \frac{\sigma^2}{n_2}$$

Therefore, when $n_1 = n_2 > n_3, V(\bar{Y}) < V(\bar{Y}_1)$

$< V(\bar{Y}_2)$

Case 2: Let $n_1 < n_2$ and $n_1, n_2 > n_3$

If so, Let us examine whether

$$V(\bar{Y}) = \sigma^2 \frac{(n_1+n_2+2n_3)}{(n_1+n_2)^2} < V(\bar{Y}_1) = \frac{\sigma^2}{n_1} \quad ?$$

i.e. $\sigma^2(n_1+n_2+2n_3)n_1 < \sigma^2(n_1+n_2)^2$

i.e. $n_1^2+n_1n_2+2n_1n_3 < n_1^2+n_2^2+2n_1n_2$

i.e. $2n_1n_3 < n_2^2+n_1n_2 = n_2(n_1+n_2)$

i.e. $n_1n_3+n_1n_3 < n_2(n_1+n_2)$ which is true.

Therefore, when $n_1 < n_2$ and $n_1, n_2 > n_3$,

$$V(\bar{Y}) < V(\bar{Y}_1) = \frac{\sigma^2}{n_1}$$

Case 3: Let $n_1 < n_2$ and $n_1, n_2 > n_3$

Is $V(\bar{Y}) < V(\bar{Y}_2) = \dfrac{\sigma^2}{n_2}$?

i.e. $V(\bar{Y}) = \sigma^2 \dfrac{(n_1 + n_2 + 2n_3)}{(n_1 + n_2)^2} < \dfrac{\sigma^2}{n_2}$?

i.e. $(n_1 + n_2 + 2n_3)n_2 < n_1^2 + n_2^2 + 2n_1 n_2$

i.e. $2n_2 n_3 < n_1(n_1 + n_2)$

The above statement is true,

If $n_3 < \dfrac{n_1}{2}$

i.e. $2n_2 n_3 < n_1 n_2$

$< n_1(n_1 + n_2)$

i.e. $n_3 < \dfrac{n_1(n_1 + n_2)}{2n_2}$

Therefore, $V(\bar{Y}) < V(\bar{Y}_2)$ if $n_3 < \dfrac{n_1(n_1 + n_2)}{2n_2}$.

The above cases 2 and 3 thus suggest that

when $n_1 < n_2$ and $n_1, n_2 > n_3$

$V(\bar{Y}) < V(\bar{Y}_1)$ and

$V(\bar{Y}) < V(\bar{Y}_2)$, if $n_3 < \dfrac{n_1(n_1 + n_2)}{2n_2}$.

Similarly, when $n_1 > n_2$ and $n_1, n_2 > n_3$

$V(\bar{Y}) < V(\bar{Y}_2)$ and

$V(\bar{Y}) < V(\bar{Y}_1)$, if $n_3 < \dfrac{n_2(n_1 + n_2)}{2n_1}$.

The above results indicate that given the data from a series of samples drawn from the same population, the variance of the average over all occasions is always expected to be lower than that of the mean observed on any occasion, in all the situations stated above, except the last one wherein a part of the sample is retained in drawing sample from the same population on each occasion. In this situation, such a precision mainly depends on the size of the sample on each occasion and size of the matched cases in the whole sample in the repeat surveys. For example, when the samples are of same size n on each occasion (Case 1, Situation 3), the variance of the combined estimate is also noted to be lower than that of the mean observed on any occasion. On the other hand, such a precision depends on the size of matched cases, if the samples are of different sizes in the repeat surveys (Case 2 and 3, Situation 3). Thus for obtaining average over all occasions, minimum variance is obtained either by keeping the same sample or by changing it on every occasion. In other words, if we wish to maximize precision and provide the best combined estimate, by averaging over all occasions, it is best to draw a new sample on each occasion, although the total size of the sample in the repeat surveys need not be the same on each occasion. Similarly, one can retain the same sample throughout all occasions to maximize precision in estimating average over all occasions, although this procedure is equally efficient in estimating the change in the mean from one occasion to the next.

Since SRS follows basically the latter sampling frame, that is, by keeping the same sampling units on every occasion in providing yearly estimates of vital rates, an average taken over a few years for a population, in-which time changes are slow during that period, appears to be adequate in most instances. Similarly, since the national/state level surveys like NFHS follow basically the former design, that is, by changing the sample on each occasion, an average taken over all occasions may provide a better estimate, particularly in the study of mortality.

DATA

The present study has primarily used two rounds of NFHS data obtained during the last decade. Data for the NFHS-1 were collected in 1992-93 from a probability sample of 89,777 ever-married women aged 13-49 years, residing in 88,562 households, which was designed to provide statistical estimates for the country as whole, and for rural and urban areas (IIPS, Mumbai, 1995). The second National Family Health Survey (NFHS-2), conducted in 1998-99 also covers an almost equal number of independent sample, that is a nationally representative sample of 89,199 ever-married women aged 15-49, residing in 91,196 households (IIPS and ORC Macro, 2000). The NFHS is one of the most complete surveys of its kind ever conducted in India. The main objective of the NFHS-1 was to collect reliable and up-to-date information on fertility, family planning, mortality and maternal and child health (including child nutrition). Most of the types of information collected in NFHS-1 were also collected in the second round of the survey (NFHS-2), with the objective also to identify trends in various indicators over the intervening period of six and one-half years. In addition, the NFHS-2 covered a number of new or expanded topics with important policy implications, such as reproductive health, women's autonomy, domestic violence, women's nutrition, anaemia and salt iodization.

Three types of questionnaires viz. the Household Questionnaire, the Woman's Questionnaire, and the Village Questionnaire, were used in the NFHS-1 and NFHS-2 to collect information on the above aspects. The Household Questionnaire listed all usual residents in each sample household plus any visitor who had stayed in the household the night before the interview. For each listed person, the survey collected basic information on age, sex, marital status, relationship to the head of the household, education, and occupation as well as health status. In addition to other household level information including household condition, the household questionnaire of NFHS-1 also included household birth and death records wherein all the live births and deaths that took place within the last two years preceding the survey in the household were recorded. In NFHS-2, the Household Questionnaire, in addition to other information, also asked about deaths occurring to household members in the two years before the survey, with particular attention to maternal mortality. Thus, the

NFHS-1 and NFHS-2 include questions on the number of deaths occurring to usual residents in each household during a particular time period, in the household questionnaire, while in the Woman's Questionnaire a complete birth history, including sex, date of birth, survival status for each live birth, and age at death for those children who had died, were included for detailed analysis of fertility and mortality in the surveyed population. As mentioned earlier, Office of the Registrar General of India has been publishing estimates of age-specific death rates every year, based on household birth and death records under the SRS, which were also used for the present study.

METHOD AND ANALYTICAL FRAMEWORK

The age specific death rates (ASDRs) along with their sampling error have been derived, based on deaths occurring to usual residents of the household during the two years preceding the survey, as obtained from the Household Questionnaire of NFHS-1 and NFHS-2. These estimates of NFHS-1 and NFHS-2 which would approximately refer to the period 1991-92 and 1997-98 respectively, have been compared with that of the combined estimate, the over-all mean for the two occasions, along with their corresponding sampling error to establish efficacy of the combined estimate. The ASDRs and their sampling errors have been calculated in the present study through the Integrated Micro-Computer Processing System (IMPS 4.1), which is used for data entry, editing, tabulation, management and dissemination of census and survey data (U. S. Census Bureau, 1998).

It may however be noted that questions on the number of deaths occurring to usual residents in each household during a reference period have usually been included as a matter of practice in demographic surveys in many countries, and have generally resulted in a substantial underreporting of deaths (IIPS, 1995, IIPS & ORC MACRO, 2000). However, in the absence of reliable civil registration data in a developing country like India, one needs to rely on such demographic survey data to derive the age-sex specific mortality pattern for the population. Nevertheless, since the Sample Registration System (SRS) is also a large-scale demographic sample survey based on the mechanism of a dual record system with the objective of providing reliable estimate of fertility and mortality indicators at state and national levels, the combined estimates of NFHS-1 and NFHS-2 will be compared with that of the SRS for the corresponding period to understand the extent of completeness of reporting of deaths. In the light of this assessment, the combined estimate of age specific death rates along with its sampling error will provide a useful base to study mortality pattern in the country. After having examined the death rates by age and sex, these data will further be used as input to construct sex specific abridged life tables for the last decade for India and this would be compared with SRS based life tables for the same period.

Construction of Life Tables

A life table is designed essentially to measure mortality and is used by public health workers, demographers, actuaries, and many others in studies of longevity, fertility, migration and population growth, as well as in making projections of population size and characteristics and in studies of widowhood, orphan-hood, length of married life, length of working life and length of disability-free life. Life tables are, in essence, one form of combining mortality rates of a population at different ages into a single statistical model. In other words, they are principally used to measure the level of mortality of the population involved. One of their main advantages over other measures of mortality is that they do not reflect the effects of age distribution of an actual population and do not require the adoption of a standard population for acceptable comparisons of levels of mortality in different populations or in the same population over time.

There are basically two types of life tables according to the reference year of the table – (i) the current or period life table and (ii) the generation or cohort life table. The first type of table is based on the experience over a short period of time, such as a year, three years or an intercensal period, in which mortality has remained substantially the same. This type of table represents the combined experience by age of the population in a particular short period of time, and it does not represent the mortality experience of an actual cohort. It assumes a hypothetical cohort that is subjected to the age specific death rates observed in the particular period. It is therefore an excellent summary measure of mortality in a year or a short period.

The second type of life table, the generation life table, is based on the mortality rates experienced by a particular birth cohort, that is, all persons born in a particular year. In other words, the mortality experience of the persons in the cohort would be observed from their moment of birth through each consecutive age in the successive calendar years until all of them die. Obviously, data over a long period of years are needed to complete a single table. Thus, the generational life tables are based on probability estimates of proportion surviving to a given age from an initial sample cohort and it is possible to construct generation life tables for cohorts born in this century, while the cross-sectional life tables are obtained from estimates of age-specific mortality rates from sample population in different age groups. As such, expected values and variances of several life table functions were obtained earlier for both the generational and the cross-sectional life tables (Wilson, 1938; Chiang, 1980; Keyfitz, 1968; Mitra, 1972). Obviously, the latter method is more realistic and can be put into use in cases where life table functions cannot be derived for the total population. Thus, the latter method will be used to derive the expected values of the various life table functions, which is based on the transformation of estimated age-specific mortality rates to probabilities of dying. It may be noted that these methods are particularly useful for countries where national vital statistics are not adequate for construction of life tables, although relevant mortality data are available from the national level sample survey.

Life tables are also classified into two types – complete and abridged – according to the length of the age interval in which the data are presented. A complete life table contains data for every single year of age from birth to the last applicable age, while an abridged life table contains data by intervals of 5 or 10 years of age. The simpler abridged life table is usually prepared than the more elaborate complete life table, as values for 5 or 10 years intervals are sufficiently accurate for most purposes and the abridged life table is less burdensome to prepare. Moreover, it is often more convenient to use in most of the situations. Therefore, the emphasis will be on the construction of abridged life tables, obviously referring to only the current life table.

A life table is constructed on the basis of certain assumptions, which can be seen to be mathematical simplifications of real life situations. As you are aware, a life-table population is considered as both stationary and closed to migration (A stationary population is defined as a population whose total number and distribution by age do not change with time). Secondly, the life table population is the total of the pieces of several cohorts, who being born at different times should have been exposed to different conditions. But once the life table is constructed the population is treated as a single cohort. Thirdly, age-specific death schedule is also assumed to operate in the set pattern and periodic variation due to causes random or otherwise is not anticipated. The other assumptions are discussed in defining the various columns of the life table.

Life Table Functions

The basic life table functions which are generally calculated and published for every life table, are ${}_nq_x$, l_x, ${}_nd_x$, ${}_nL_x$, T_x and $\overset{o}{e}_x$, although in some cases, due to limitations of space, some of these columns may be omitted. In general, the mortality rate, ${}_nq_x$, is the most basic function in the table, that is, the initial function from which all other life table functions are derived. The life table model conceptually traces a cohort of newborn babies through their entire life under the assumption that they are subjected to the current observed schedules of age-specific mortality rates. The cohort of newborn babies, called the radix of the table, is usually assumed to number 100,000. In this case, the interpretation of the life table functions in an abridged table would be as follows:

i. Age x to x+n: The first column of the life table is age and is represented by x. Age x to x+n is the age interval, indicating the period of life between two exact ages.

ii. ${}_nq_x$: The proportion of the persons in the cohort alive at the beginning of an indicated age interval (x) who will die reaching the end of that age interval (x+n). In other words, it indicates mortality rate, that is, the probability that a person at his/her x^{th} birthday will die before reaching his/her $x+n^{th}$ birthday.

iii. l_x : The number of persons living at the beginning of the indicated age interval (x) out of the total number of births assumed as the radix of the table. Thus, l_0 is the size of the birth cohort and l_x is the number of survivors at exact age x.

iv. ${}_nd_x$: The number of deaths, that is, the number of persons who would die within the indicated age interval (x to x+n) out of the total number of births assumed in the table. In other words, ${}_nd_x = l_x \cdot {}_nq_x$.

v. ${}_nL_x$: The number of persons years that would be lived within the indicated age interval (x to x+n) by the cohort of 100,000 births assumed. In other words, the number-years lived by the l_x persons during the interval (x to x+n), which is equivalent of the population and therefore is called the life table population.

vi. T_x : The total number of person-years that would be lived after the beginning of the indicated age interval by the cohort of 100,000 births assumed. In other words, the value of T_x for any x is defined by cumulating ${}_nL_x$ from x to w, that is, $T_x = \sum_{h=0}^{w} nL_{x+nh}$

vii. $\overset{o}{e}_x$: The last column of the life table gives the average remaining life time (in years) for a person who survives to the beginning of the indicated age interval. This function is also called the complete expectation of life, or simply life expectancy at age x and is conventionally denoted as $\overset{o}{e}_x$. It is computed by dividing column T_x by l_x, that is $\overset{o}{e}_x = \frac{T_x}{l_x}$

Method of Constructing Abridged Life Table

As discussed earlier, the most fundamental step in life table construction is to obtain the values of ${}_nq_x$ the mortality rates or age-specific probabilities of dying by converting the observed age-specific death rates with the assumption that $m_x = M_x$ where m_x is the central age-specific life table death rates and M_x is the age-specific death rates of the population. In a complete life table, the basic formula for converting the observed age-specific death rates into their corresponding age-specific probabilities of dying is

$$q_x = \frac{2m_x}{2+m_x}$$

where m_x is the observed death rate at a given age and q_x is the corresponding probability of dying. This formula is based on the assumption that deaths between exact ages x and x+1 are uniformly distributed, which in other words, implies that l_x is linear.

There are a number of methods suggested for the construction of abridged life tables (King, 1912; Reed and Merrell, 1939; Greville, 1943; Chiang, 1968). One of the key features of the various short-cut methods is the procedure for making this basic transformation of m_x values into q_x values, from which other columns of the life table can be derived. Another difference in the method is in the way the stationary population, L_x, is derived. In this context, a method suggested by Greville (1943) is described here as this method was followed for the present study to construct the abridged life tables. This is one of the short-cut methods and is simple in estimating ${}_nq_x$ and ${}_nL_x$ values. The observed age-specific death rates are converted to the needed mortality rates by the use of the formula:

$$ {}_nq_x = \frac{{}_nm_x}{\frac{1}{n} + {}_nm_x[\frac{1}{2} + \frac{n}{12}({}_nm_x - \log_e c)]} \sum_{i=1}^{n} X_i Y_i $$

In this equation, 'c' comes from an assumption that the ${}_nm_x$ values follow an exponential curve (Shryock et.al., 1976). Empirically, the value of c was estimated to be between 1.08 and 1.10, and therefore $Log_e c$ was assumed to be about .095 as an intermediate value. The use of modern computers has greatly reduced the work of derivation of ${}_nq_x$ values by this method and the standard software package is available for direct calculation by electronic computer on the basis of ${}_nm_x$ values and two constants n and $Log_e c$.

In Greville's method, the central death rates in the life table and the population are assumed to be the same, and the desired value of ${}_nL_x$ is derived by the use of the formula

$$_nL_x = \frac{nd_x}{nm_x}$$

For the last age interval, that is, the interval with the indefinite upper age limit, the value of $_nL_x$ is approximately estimated by $_nL_x = \frac{l_x}{q^{m_x}}$.

As mentioned earlier, the combined age-specific death rates, estimated from the two rounds of NFHS conducted during the last decade were used as input to construct the abridged life tables for India by this method, using the software package for mortality measurement viz. MORTPAK 4.0 for Windows, developed by United Nations Population Division (2003).

Standard Error of the Estimate of Life Expectancy

The present study also attempts to calculate life expectancy at birth and at higher ages with 95 percent confidence intervals for sub-national areas in the country. The method by Chiang (1978; 1984) allows the calculation of the standard error even when there are zero deaths present in age bands, because the calculation of $Var(q_x)$ involves multiplying by the mortality rate, and has been successfully used by other studies

(http://www.statistics.gov.uk/statbase/product.asp).

The variance of life expectancy at age x (e_x^o) is obtained by

$$Var\left(e_x^o\right) = \frac{\Sigma\left[l_x^2\left\{\left(1-a_x\right)n+e_{x+n}\right\}^2 .Var\left(q_x\right)^2\right]}{l_x^2}$$

The standard error of e_x^o is given by

$$SE\left(e_x^o\right) = \sqrt{Var\left(e_x^o\right)}$$

Where, $Var(q_x) = \dfrac{n^2.M_x[1 - a_x.n.M_x]}{P_x[1 + (1 - a_x)n.M_x]^3}$

Where,

q_x = Conditional probability that an individual who has survived to start of the age interval, will die in the age interval.

a_x = Fraction of the age interval lived by those in the cohort population who die in the interval.

P_x = Population in the age interval.

M_x = Age specific death rate.

The methods outlined by Chiang (1984) were therefore used here to calculate the standard error of the life expectancy at different ages, the results of which are presented and discussed in the relevant section of various chapters.