CHAPTER-2

DATA AND METHODS
2.1 Introduction

The most precious incidents of life for any human being are birth, marriage and death. These are known as vital events. In India, recording of vital events was started prior to the first census year 1872. One type of data cannot fulfill all demands of the government and policy makers. Presently, principal sources of vital data are censuses, civil registration system (CRS), sample registration system (SRS), demographic sample surveys like national sample survey (NSS), National family health survey (NFHS), District level household and facility survey (DLHS), etc. Here, the matter of choice is only data related to death.

In the subsequent section, thesis critically assessed the benefit and limitation of different sources of mortality data through table 2.1 which is borrowed from a study done by Mahapatra (2010). Table 2.1 gives an overview of all above mentioned sources of vital statistics in India. A census is the procedure of systematically getting and recording information about the each member of a given population (Alterman and Hyman, 1969). In India, the census is the most convincing source of information on population characteristics, economic activity, literacy, household amenities, fertility and mortality and many other demographic data since 1872 (MOSPI, 2010). In order to conduct a census, enumerator questions one or more respondents in each household to obtain information about the household, its member, and the dwelling unit they occupy.
In India, the census is conducted for every ten years. However, country needs to know the number of births and deaths each year and the main causes of their deaths for better development and operation of the health system. Thus, the government started to count everyone and to track all births and deaths through civil registration system (CRS).

Table 2.1: An overview of sources of vital statistics in India.

<table>
<thead>
<tr>
<th>Source</th>
<th>Periodicity</th>
<th>Estimated Parameters</th>
<th>Small-Area Parameters</th>
<th>Usability</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRS</td>
<td>Annual, Since 1958</td>
<td>Fertility and Mortality Indicators</td>
<td>District Level and Large Cities with more than 100000 population</td>
<td>Less than 50% of deaths registered. Wide interstate variation. The Average time to publication: 45 months until 1994. No report since then.</td>
</tr>
</tbody>
</table>

\(^a\) SRS: Sample Registration System, Operated by the Registrar General, India  
\(^b\) CRS: Civil Registration System, Operated by Local Bodies, Managed by State Governments, Tabulation, Publication and National Coordination by the Registrar General, India  
\(^c\) NFHS: National Family Health Survey, Operated By the International Institute of Population Sciences, Mumbai  
\(^d\) DLHS: District Level Household and Facility Survey, Operated By the International Institute of Population Sciences, Mumbai

Source: Mahapatra, P. (2010) and further updated by the researcher
CRS provides the basis for individual legal identity but also allows countries to identify their most pressing health issues (WHO, 2013). Although, the civil registry is crucial tool for gathering information on vital statistics, it has many limitations and disadvantages. Primary disadvantages are; very costly to establish, to maintain and incompleteness because of non-reporting, etc.

In the civil registration system, there was no uniformity in collected information in the context of both registration and coverage. Thus, in order to unify the civil registration system and with a view to generate reliable and continuous vital data, office of the Registrar General of India (ORGI) initiated a scheme, called as sample registration system (SRS). It started in 1964-65 on a pilot basis and outset on a full scale from 1969-70. The SRS since then has been providing vital data on a regular basis. It is based on a system of dual recording of births and deaths in reasonably representative sample unit spread all over the country. The SRS provides annual estimates of population composition, fertility, mortality and medical attention at the time of birth or death which gives some idea of access to medical care.

National Sample Survey (NSS) provides reliable and regular information on various issues such as employment, morbidity, health care, expenditure, education. NSS conducted through household interviews, using a random sample of households covering the entire geographical area of the country practically. National Family Health Survey (NFHS, 1992-93, 1998-99 and 2005-06) furnishes health and demographic data on fertility, child and infant mortality, morbidities, health care services, etc. (IIPS, 1995; IIPS and ORC Macro, 2000; IIPS and
Macro International, 2007). Among the above-mentioned sources, SRS is the most reliable and regular sources of vital data for the country and major States having a population larger than ten million (Mahapatra, 2010).

This study uses the secondary data on age-specific death rate (ASDR), crude death rate (CDR), and infant mortality rate (IMR) from SRS to examine the level, trend and pattern of mortality and to construct a life table. The Study also exploited the data from the report of the publication of the Office of the Registrar General of India “District Level Estimates of Child Mortality in India based on Census 2001 data.” (See ORGI, 2009)

For the purpose, thesis considered all 17 major states of India for which mortality data is available from 1970s. For first and second objective, the study analyzed data for only seventeen major states of India for which death information is available in SRS. For all major states, except Bihar and West Bengal, life tables are available from 1970s onwards. Nevertheless, for Bihar and West Bengal life tables are in use from 1980s. The Union Territories were not included in the analysis because of data constraints. The state level study has been done by sex (i.e. male and female) and place of residence (i.e. rural and urban). The two distinct sources of data used in the study, to examine the changing age pattern of mortality and construction of the district level model based life table, are described in detail in the next section.
2.2 Data Sources

In this section, data sources used in the study are described in detail.

2.2.1 Sample registration system

The changing dynamics of Indian population and the deficiency in civil registration system, in terms of under registration and incompleteness in coverage, have mostly indicated the need for quick and reliable estimates of births and death rates on a current and continuous basis. Though, the Indian census provides decadal information; but not able to deliver the changes in vital rates across the year. Therefore, to assess the short-term changes in growth of population, Office of Registrar General of India (ORGI), initiated a scheme of sample registration of birth and death data. Initially in 1964-65, registration of births and deaths through SRS started on a pilot basis only in rural areas and was implemented on a full scale in 1969-70. Since then SRS is providing vital data on a regular basis.

The primary objective of SRS is to render reliable annual estimates of birth and death rates at the state and national level for rural and urban areas as well as for each sex separately. It also supplies the various measures of fertility and mortality. It is based on the dual record system and having quite a good representative sample units spread all over the country. The system calls for continuous registration by a designated local registrar and half-yearly survey by a supervisor. The events then matched; unmatched events are verified and included only after confirmation (Kulkarni, 2011). The sampling design, registration and validation of the methodology of SRS has been published (ORGI, 1972; ORGI, 1993). The sampling frame get revised every ten years on the basis of the latest census frame and old sample clusters are
replaced by new ones (Mahapatra, 2010). The sample unit in rural areas is a village or a segment of it if the village population is 2000 or more. In urban areas, the sampling unit is a census enumeration block (CEB) with a population ranging from 750 to 1000. At present, SRS is operational in 7,597 sample units (4,433 rural and 3,164 urban) spread across all states and union territories and covers about 1.5 million households and 7.27 million population (ORGI, 2011).

SRS is the only source in India, which renders the age-specific death rates (ASDRs) for each year on a regular basis. It is a vital source of information on birth and death rates. SRS assumed as most reliable source of information on ASDRs, birth rates and infant mortality rate by sex and place of residence in India and major states. Therefore, study chooses the SRS to access the age pattern of mortality in India. It provides the ASDRs from 1970 to 2010 by quinquennial age group for bigger states of India. However, for state Bihar and West Bengal, SRS provides the ASDRs from year 1981 onwards. In the similar way, the ASDR data for state Jammu & Kashmir is found irregular. The data is available for the period 1971 to 1990 and then for 2004 onwards regularly collected. Along with, the state Uttar Pradesh, Madhya Pradesh and Bihar carved into Uttarakhand, Chhattisgarh and Jharkhand respectively in year 2000. So, the life table of the state Uttar Pradesh, Madhya Pradesh and Bihar from year 2000 onwards refers to only those parent states; it does not include mortality information of Uttarakhand, Chhattisgarh and Jharkhand respectively.
The ASDRs is available up to the age 70+ over the period 1970 to 1994 and up to 85+ for the period 1995 to 2010. SRS furnishes ASDRs for 0-4 age group, for the period 1970 to 1995 and for the period 1996 to 2010 data is available for ages 0-1 and 1-4 separately.

2.2.1.1 Quality of SRS data

The quality of any data can be accessed through coverage, completeness and incidence of missing in terms of non-reporting, unavailability and wrong entry. The evaluation of SRS data has been made several times, both in the house by the Registrar General of India (RGI) and other authors. Though, it is believed that Sample Registration System of India provides reliable estimates of birth rate and death rate at the national level and state level as well, it also suffers from incompleteness in birth and death registration. Direct and indirect evaluation both showed that the incidence of under-registration of births and deaths were within the tolerable range up to 10% (Mahapatra, 2010). However, most of the studies based on evaluations of SRS were conducted during 1970s and 1980s. Brass in a study of 1975 indicated, 6% under-reporting of adult deaths (ORGI, 1982). Preston and Coale in 1980 pointed out 10% under-reporting of deaths in SRS (Bhatt et. al., 1984). Rele and Plamore in 1992 in research showed that SRS may be overestimating mortality rates (Swamy, et. al., 1992). Mahapatra (2010) conveyed that the registration of deaths within sample areas, by SRS, has worsened during 1990s up to until 2007. However, the completeness at national level ranges between 77 to 99 percent. He also stated that the interstate differences in the completeness of SRS data emerges widened.
Some of the other possible deficiencies outlined here:

- SRS is failure to record all vital events as they occur.
- Children who show signs of life after birth but die within few seconds/minutes or afterward, possibly reported as stillbirths. SRS can suffer from this bias that results in an underestimation of mortality.
- SRS is not providing the estimates below state level i.e. district level, as well as an estimate of LEB is not available for smaller states.
- Age heaping involved in reporting of mortality data also leads the inadequacy in data.

2.2.1.2 Assumption

On the whole, the Indian SRS data has been considered as reliable and trusted source of fertility and mortality statistics at national and major states level. Though, SRS covers only about 0.6 % of India’s population, but its representative characters allow for estimation of vital data for the country and major states (Mahapatra, 2010). In addition, consistency over time in definitions of terms, administrative procedure and method used for data collection, permit the comparability across a period and over country. Therefore, in this regard, study assumes that Sample Registration system (SRS) based ASDRs of major States and India of the single years 1970 to 2010 reported by the Registrar General of India are good enough to construct model life tables.


2.2.2 Census of India

It is conducted by the Office of Registrar General and Census Commissioner, India under the Ministry of Home Affairs, Government of India. As stated earlier, Census covers the various aspects of population, economy, socio-cultural aspects, migration area and village profile, etc. For the present research, study needed information on either ASDRs or infant/child mortality at the district level, but no surveys gives district wise ASDRs. Therefore, the study started to look for the estimates of infant and child mortality for all districts of India. Ideally, the civil registration system should be able to provide these estimates regularly. However, because of high coverage error and incompleteness, civil registration is not satisfactory.

The SRS has been rendering the estimates of infant mortality over the four decades at the state level. Some recent SRS reports also give the estimates for natural regions in large states, but for the requirement of planning and policies below the state level, it is desirable to have district level estimates. The National Family Health Survey (NFHS) also allows for infant and childhood mortality at the state level, not at the district level. Some other surveys like District level Household and Facility Survey (DLHS) in India collect data of child survival, but the sample size at the district level is not enough to produce the estimates with tolerable sampling error. Therefore, the demand is realized to supply the infant and child mortality estimates at the district level.

In this regard, Indian census included some questions, like number and sex of children ever born and number and sex of children surviving, in 1981 and computed the district level

The present study is using the information on IMR derived from Census 2001. The information on IMR has been collected from the publication of the Office of the Registrar General of India “District Level Estimates of Child Mortality in India based on Census 2001 data.” In this report, IMR is indirectly estimated using Brass technique. IMR is available sex and place of residence wise for all 593 districts from state and union territories of India (ORGI, 2009).

### 2.3 Method of analysis

#### 2.3.1 The life Table: A Demographic Overview

A life table is designed fundamentally to quantify mortality. It 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, orphanhood, length of married life, length of working life and length of disability-free life (Shryock et. al., 1976). Life tables are in the real meaning, one form of combining mortality rates of the population at different ages into a single statistical model (Das and Patel, 2004).
Life tables are of two types. The first type is known as current or period life table. This type of life table is based on the experience over a short period viz. a year, three years or an intercensal period. These tables represent the combined experience by the age of the population in a particular short time and do not correspond the mortality experience of an actual cohort. It assumes a hypothetical cohort that subjected to the age-specific death rates observed in a particular period.

The second type of life table, the generation (cohort) life table, is based on the mortality rates experienced by a particular birth cohort, that is, all person born in a particular year. In other words, the mortality experience of the individuals 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. Clearly, data over a long period of years are needed to complete a single table.

Life tables also categorized into two types – complete and abridged – according to the length of the age interval in which the data are presented. A complete life table includes data for every single year of age from birth to the last applicable age while an abridged life table contains data at intervals of 5 or 10 years of age. In the thesis, the focus is on the construction of abridged life table referring only the current life table for states and districts of India.

A life table is created on the basis of certain assumptions, which can be considered to be mathematical simplifications of real life situations. These are; first a life-table population is considered as both stationary and closed to migration. Here it is to clarify, a stationary
population defined as a population whose total number and distribution by age do not change with time. Secondly, the life table population is 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 constructed, the population is treated as a single cohort. Thirdly, age-specific death schedule is also assumed to operate in a set pattern and periodic variation due to causes random or otherwise not anticipated (Shryock et. al., 1976). The other assumptions discussed in defining the various columns of the life table.

2.3.1.1 Importance of Life Table in Mortality analysis

Mortality study can be done in a number of ways. Two methods are commonly in used. In one way, crude and age-specific death rates are considered and in another, life table idea is used (Kohli, 1977). The main limitation of crude death rates is its dependence on age and sex structure of the population as crude death rates of two populations can be similar, but their age structure may differ remarkably. On the other hand, death rates are purely a function of age-specific death rate and independent of sex and age composition of the population (Kohli, 1977). It is, therefore, more appropriate to use life table concept to perform mortality analysis. Demographers agree on the statement that the mortality pattern of the human population as described in life tables can be distinguished not only in terms of level of mortality but also in terms of distribution of deaths among various age groups (Anson, 1991).

Mortality studies primarily using life tables are not new to the Demographers and policy planners. This branch of demographic research has a history of more than three centuries. John Grant was the first, who introduced this concept in his work published in 1662 based on the
“Bills of Mortality.” Since then, many researchers have been modified and introduced life table techniques.

However, the major developments in the study of life tables have been taken place during the last 100 years. The significant developments in life table technique during the above period were carried out by King (1914), Reed and Merrell (1939), Greville (1943), Chiang (1968), . Thus, the life table is the most appropriate tool for the pattern of mortality analysis.

2.3.1.2 Anatomy of Life Table

Construction of cohort and current life tables are identical in appearance but different in structure (Chiang, 1984). The following discussion refers to complete current life table. Each column defined below:

**Column 1:** Age interval, x to x+n: Age interval between exact ages for each row of the life table except for the final age interval that is open-ended (e.g. 85+).

**Column 2:** \( nq_x \): The proportion of the population in each age interval that are alive at the beginning (at age x) of the interval and dead before reaching the end of the interval (x+n). The proportion is computed from the observed mortality rates of the actual population and is used to derive the remaining columns of the life table.

**Column 3:** \( l_x \): The number of persons alive at the beginning of the age x. The first number in this column, \( l_0 \), is an arbitrary figure called as ‘radix.’

**Column 4:** \( d_x \): The number of persons dying during the age interval (x, x+n).
**Column 5:** \( L_x \): The total number of person-years in the stationary population for each age interval. It can be viewed as the average population size between birthdays, taking into account the distribution of deaths throughout the year.

**Column 6:** \( T_x \): This column records the stationary population in the indicated age interval and all subsequent intervals. It is the cumulative sum of the \( nL_x \) values. It could be viewed as the total number of person-years that would be lived for a particular age cohort if the cohort were to progress through the remainder of the life table.

**Column 7:** \( e_x \): This is the number of years, on average, yet to be survived by a person of age \( x \).

### 2.3.1.3 Life table construction

The analysis of the present thesis built upon the information provided through the sample registration system (SRS) in this country. From the previous section, it is known that SRS provides annual estimates of ASDRs for India and major states by sex and place of residence. It is observed that ASDRs available through SRS have been associated with the some random fluctuations of unknown origin (Chaurasia, 2006). To abolish these fluctuations, Registrar General of India (1984; 1985; 1989; 1994; 1998; 2004) produced life table on the basis of ASDRs for five year period rather than annually. However, the problem associated with these life tables are that they are based on different methods of converting ASDRs into the life table function \( nq_x \). The life tables for the period 1970-75, 1976-80 and 1981-85 are based on Greville’s method (Greville, 1948). Whereas life table for period 1986-90 onwards are prepared using the Mortpak software package (United Nations, 1988) developed by the United Nations (Chaurasia, 2006). MORTPAK also uses the Greville method. However, some o
formulae used here differ from the one used in deriving the earlier life tables namely 1970-75, 1976-80 and 1981-85.

Thus, in order to improve the similarity and consistency in estimates of life tables for different five-year period, the present researcher reconstructed the life tables for all periods, starting from 1970-75 to 2006-10, using the application LIFTB of the MORTPAK Software Package. For the input of MORTPAK in the construction of the life table, study used the unadjusted ASDRs as it was unconditionally assumed that the quality of SRS data is good enough to construct life tables. Furthermore, it is to state that the last open-ended age/age group considered in completing the life tables of all states and India of all time periods of the reconstructed life table is 85+. However, it is possible to end a life table at age 100+ using MORTPAK, but the researcher preferred to stop by age 85+ for obvious reasons.

2.3.2 Decomposition of Life Expectancy at Birth

In Chapter 3 of the thesis, the researcher employed a decomposition approach to study the changes in overall mortality from all ages over the period of India and also different from states. The method is used to see the contribution of different age mortality like infant mortality, child mortality, adult mortality and old age mortality in life expectancy at birth (LEB). According to Candus (2003) the decomposition approach is extensively used in demography in assessing the changes in a number of demographic variables over time. Several experts suggested various methods to decompose the changes in life expectancy at birth overtime into the contribution by different ages (Chandraseker, 1949; Kitagawa, 1955;
Rutherford, 1972; Andreev, 1982; Pollard, 1982; United Nations, 1982; Arriaga, 1984; United Nations 1985; Candus, 2003). Ponnapalli (2005) in a study compared some of the above decomposition approaches and concluded that many of these methods are rendering the same outcomes. He further suggested the use of the UN (1985) method for its simplicity.

Following the idea suggested by Ponnapalli (2005), present study used the decomposition approach proposed by United Nations (1985) for analyzing the mortality changes over time in case of India and its major states, sex and place of residence wise. The detailed description of the United Nations decomposition approach is supplied here.

The formulae to find the contribution of the age group x to x + n for a change in life expectancy at birth \((e^0_x)\) suggested by United Nations (1985) is as follows:

\[
e^2_0 - e^1_0 = \frac{(e^2_0 - e^1_0)(I^2_1 + I^1_1)}{2} - \frac{(e^2_{x+n} - e^1_{x+n})(I^2_{x+n} + I^1_{x+n})}{2}
\]

\[\text{......... (2.1)}\]

For the open-ended age group, formula is written as:

\[
e^2_0 - e^1_0 = \frac{(e^2_0 - e^1_0)(I^2_1 + I^1_1)}{2}
\]

\[\text{............... (2.2)}\]

Where all the notations are in usual manner and described below:

\(e^0_1\) = Expectation of life at birth in the initial period 1

\(e^0_2\) = Expectation of life at birth in the later period 2

\(e^1_x\) = Expectation of life at exact age x, in the initial period 1

\(e^2_x\) = Expectation of life at exact age x, in the latter period 2
1_{1x} = Number of persons alive at exact age x, in the initial period 1

1_{2x} = Number of persons alive at exact age x, in the latter period 2

l_{10} = l_{20} = Radix of the life table in the initial and later time periods 1 and 2

n = Length of the age interval

2.3.3 Comparison of mortality models

2.3.3.1 Method for comparison of mortality models

In the present section, an effort is made to present a variety of observed sex and age pattern of mortality and method to investigate the applicability of the existing empirical mortality models to Indian mortality data. However, there is no standard procedure for testing the closeness of mortality patterns in hand to any model life table. Since, the relationship between infant mortality, child mortality and adult mortality is crucial in determining the selection of the model (Parasuraman, 1990), it is used as a criterion for judging the applicability of models. It is noticed that mortality below age 10 is high, and their levels significantly vary among the states of India. There are many states with different levels of infant and child mortality but with identical adult mortality levels (e^{0.10}) (Parasuraman, 1990). Thus, the expectation of life at age 10 (e^{0.10}), is selected as a measure of adult mortality. Corresponding to the level of e^{0.10}, life table functions like q_0, q_1 and, q_5 are estimated from each of the nine patterns in the United Nations and Coale & Demeny life table for India and major states sex wise. After the estimation models giving the best results for Indian mortality data, are identified.
The functions of model life tables (e.g. \( q_0 \), \( 4q_1 \) and, \( 5q_5 \)) corresponding to actual life tables for United Nations and Coale & Demeny model life table are obtained by the application of ‘COMPAR’ program which is a software package available in MORTPAK developed by United Nations (United Nations, 2003). The description of this program is given as follows:

**Purpose of COMPAR program:** It compares empirical set of age-specific central death rates \( \mu_m \) or age-specific probabilities of dying \( \mu_q \) values to all United Nations and Coale-Demeny model life table patterns and publishes indices of similarity too.

**Description of technique:** For each age-specific mortality rates either \( \mu_m \) or \( \mu_q \) values are given as input. Corresponding life expectancy at birth in each of the five United Nations pattern (Latin American, Chilean, South Asian, Far Eastern and General Pattern) and four Coale-Demeny models (North, South, East and West) is detected by calling another MORTPAK program known as MATCH. This program calculates the United Nations, Coale and Demeny or user-designated model life tables corresponding to the given levels of mortality. As the user-designated model can be a mortality pattern specific to a certain population, MATCH can generate a country-specific model life table system. For each model, a series of life expectancies that are more or less constant by age indicates that the empirical mortality pattern is similar to that model. This program also produces the indices of the goodness of fit for age group 0-10, 10 & above and 0 & above.

In addition, the Indian mortality pattern is also compared with WHO model life tables which is also known as the Modified Brass Logit System (MBL). However, the function (e.g. \( q_0 \), \( 4q_1 \)
and, $5q_5$ of modified Brass logit model life table is calculated using the ‘modmatch’ command available in software STATA (StataCorp, 2011). In this model, one must define sex of the life table, the value of $a_1$, (It is the average number of years lived in the age interval $(0, 1)$ by those dying in the same age interval $(0, 1)$) and two life table functions as input. At the same time, one input must be $l_5$ or $5q_5$, whereas the second input may be $45q_{15}$ or $e_0$ or just left blank. This software will automatically estimate the value of $l_{60}$ from $l_5$ based on the observed relationship between these two parameters in the life table used to create the modified Logit system. All estimates are obtained by sex and for time periods 1971-75, 1981-85, 1991-95, 2001-2005 and 2006-10.

2.3.3.2 Examination of three Models

There is a wide variety of analytical methods to identify a set of model life tables best fit to the actual life table under study. Adlakha (1972) tested the applicability of model life tables to less developed countries. In order to check the applicability, he compared the estimated and expected values of life of expectancy at age 5 and the probability of dying before age one from the model life tables. To evaluate the observed value with model life table values, Parasuraman (1990), used the indices of goodness of fit such as mean absolute deviation from the median of life expectancy values for the age group 0 to 10, 10 and 0 & above age groups. Parasuraman also applied one more index say the difference between the medians within the age range 0 to 10 and the median for 10 and over age groups. A small value of this difference stands for consistency in the pattern. In the same way, for analyzing the mortality models and their forecasting ability with particular reference to the country Taiwan, Wang (2010) used the five measures say mean error, mean absolute error, mean square error, mean percentage
error, mean absolute percentage error. Bayesian Information Criterion (BIC) can also be used to judge the applicability of the model and statistical significant. This process also allows the comparison of prediction capabilities across models (Cavanaugh, 2012).

Initially, the study tried for both error measures and BIC approach; noticed that both are delivering almost similar results. Therefore, decided to go for the simplest method i.e. error measure method. Thus, for ensuring the goodness of fit of all the fitted models, thesis utilized three measures say mean error (ME), mean absolute error (MAE) and mean square error (MSE). Error measures play an important role in calibrating or refining a model so that it will forecast accurately for a set of time series. However, it is often difficult to obtain the appropriate measure.

The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures accuracy for continuous variables. The MAE is a linear score which means that all the individual differences are weighted equally in the average, whereas the RMSE is a quadratic scoring rule that measures the average magnitude of the error. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. It means the RMSE is more useful when large errors are particularly undesirable (Wilmott and Matsuura, 2005). Thus, the previous discussion suggests that it is difficult to choose the suitable measure for comparison.

Wilmott and Matsuura (2005), in this context viewed for the advantage of mean absolute error over the root mean square error in assessing the average model performance and found that
MAE is more natural measure than the MSE. The mathematical formula for each error measure is given below:

\[
\text{ME}(\overline{\mu_k}) = \frac{1}{n} \sum_{i=1}^{n} (\overline{\mu_k} - \overline{\mu_k})
\]  \hspace{1cm} \text{…………… (2.3)}

\[
\text{MAE}(\overline{\mu_k}) = \frac{1}{n} \sum_{i=1}^{n} |\mu_k - \overline{\mu_k}|
\]  \hspace{1cm} \text{…………… (2.4)}

\[
\text{MSE}(\overline{\mu_k}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\mu_k - \overline{\mu_k})^2}
\]  \hspace{1cm} \text{…………… (2.5)}

Where, \( \overline{\mu_k} \) is predicted value of deaths rate at age \( x \) at time \( t \); \( \mu_k \) is the true or observed value of death rate at age \( x \) at time \( t \); \( n \) is the number of observations and \( t \) is time.

Above error measures are used to test the fitness of the model for particular age-specific mortality; the use of multiple measures can be unmanageable; thus, a single error measure is desirable. To compare the complete mortality age pattern, a single index may be derived out from a set of these mortality rates that are named as “Index of relative difference (IRD).” The Index of Relative difference (IRD) can be calculated as follows:

\[
\text{IRD} = \frac{100}{2N} \left| \sum \left( \frac{n\hat{q}_k}{\hat{q}_k} - 1 \right) \right|
\]  \hspace{1cm} \text{………………………… (2.6)}

Where \( n\hat{q}_k \) = values from actual life tables

\( \hat{q}_k \) = values from model life tables
These ratios cover values of mortality rates ($nq_x$) from age 0 to 80 years comprising the set of 18 figures so the value of $N=18$ in the above formula. The value of IRD serves as a single indicator measuring the difference in two sets of age-related variables such as age structure and age-specific mortality rates (Shryock & Siegel, 1976). The smaller value of IRD shows the greater similarity between two sets of variables compared.

### 2.3.4 Regression-based model life tables

Ideally, model life table system should have some essential characteristics. First, the system should be parsimonious and call for only one or few parameters to create a full life table. Second, it should sufficiently and adequately capture the broad range of mortality age pattern observed in the actual population and must imply high predictive validity. Last, it should render an acceptable estimate of age-specific death rates for countries having high levels of mortality (Wang et. al., 2013).

Very recently, some attempts have been made by a number of researchers to develop model life tables (MLTs) using the only information on either infant or child mortality or life expectancy at age $x$ (LE($x$)) values (Sinha and Gupta, 1979; Ponnapalli, 2010; and Wilmoth, 2009). So, following this notion, the study developed a set of regression-based model life table systems, for males and females separately and both sexes also combined, for India and each of the major state.
2.3.4.1 Method for Construction of Regression based model life tables

In total, 1242 abridged life tables of male, female and both sexes combined were constructed by the researcher to use as an input in deriving the various model life tables. Each regression-based model life table system consists of 19 regression equations corresponding to each age group 0-1, 1-4, 5-9, ……., 80-84 and 85+.

First, to estimate the life expectancy at birth (LEB) using the information on infant mortality rate (IMR), study developed a regression model by taking the infant mortality rate as independent variable and life expectancy at birth as dependent or outcome variable.

This regression model has the following form:

\[
\ln(LEB) = a + b \times IMR
\]

………………… (2.7)

After estimating the life expectancy at birth (LEB or LE(0)) from above model and following the idea recommended by Gabriel and Ronen (1958) and further used by Sinha and Gupta (1979) and Ponnapalli (2010), researcher derived the remaining life expectancy values of LE(x) using the regression model of the form:

\[
\ln[LE(x)] = a + b \times \ln[LEB]
\]

………………… (2.8)

Where,

\[
a = \text{constant}
\]

\[
b = \text{regression coefficient}
\]

IMR = Infant mortality rate per live births

LE(x) = Life expectancy at age x, \(x\neq 0, x=1, 5, 10, \ldots, 85+

Ln= \text{Natural logarithm}
The R^2 values also supplied next to each regression model that explains the acceptability of the model. The models are developed for India and major states by sex as well as both sexes together. After getting the complete LE(x) column at first from equation 2.8, the full life table can be derived in reverse order by applying the following steps described below:

**Step 1:** Assume \( l_0 = 1 \). Estimate \( l_x \) column as follows:
\[
l_1 = l_0 \times \left[ 1 - \frac{1 + e_1 - e_0}{1 + e_1 - e_0} \right] \quad \text{where} \quad a_0 = 0.1
\]
\[
l_5 = l_1 \times \left[ 1 - \frac{1 + e_5 - e_1}{1 + e_5 - 4a_1} \right] \quad \text{where} \quad a_1 = 1.4
\]
\[
l_{x+n} = l_x \times \left[ 1 - \frac{1 + e_{x+n} - e_x}{1 + e_{x+n} - 5a_x} \right] \quad \text{for} \quad x = 5, 10, 15, \ldots, 85+ \quad \text{and} \quad 5a_x = 2.5
\]

**Step 2:** Calculate \( m_x = \frac{l_x - l_{x+n}}{(n \times l_x + n \times a_x) \times (l_x - l_{x+n})} \) and \( q_x = \frac{n \times m_x}{1 + (n - n \times a_x) \times m_x} \)

**Step 3:** Estimate \( d_x = l_x - l_{x+n} \)

**Step 4:** Compute \( L_x = n \times l_x + n \times a_x \times (l_x - l_{x+n}) \)

For open ended age group say 85+: \( L_{85+} = \frac{l_{85+}}{m_{85+}} \)

**Step 5:** Calculate \( T_x = T_{x+n} + nL_x \)
For open ended age group say 85+: \( T_{85+} = L_{85+} \)
2.3.4.2 Extension of new life table up to age 100+

As stated earlier, the reported ASDRs are available up to the age 70+ for the period 1970 to 1994 and up to 85+ for the period 1995 to 2010. Thus, at the first step study extended the life table of the period 1970-1995 up to age 85+ using the program LIFTB that is available in software package MORTPAK. This program uses the Graville’s method to extrapolate the ASDRs value for outside the range. Thus, state-based regression model life table when used to derive the district life table using the only information on IMR of the district provide a model based district level life table that ends at age 85+, as the model life table of the state under consideration also provides LE(x) up to age 85. However, in the recent past as population aging also become a cause of concern for the government of India and its various States especially states of the South India at present, there seems a need to have district level life tables up to age 100+. Keeping this in mind, an attempt is made to provide district level life tables up to age 100+ using a methodology earlier suggested by Coale and Guo (1989).

Coale and Guo (1989) have used their methodology for the estimation of life expectancy at older ages with an open interval above 100. (i.e. 100+) of the modified Coale and Demeny life tables as the probability of dying has raised up to age 110 in recent times for various developed countries (Vaupel et. al., 1998; Candus, 2008; Kannisto, 2000). It has long been noticed that mortality rates at ages above 75 or 80 increases with age at a diminishing rate rather than at the constant Gompertz rate (Perks, 1932). Thus, Coale and Guo (1989) modified the procedure for closing out the model life tables above age 80. In this modified procedure, they make an assumption of a steady decrease rather than Gompertzian constancy in the rate of increase in mortality with age above 80. The method required the age-specific mortality
rate \( (s_{mx}) \) as input for age 75-80 and 80-85. Therefore, to compute the mortality rates at older ages they suggested the following steps:

**Step 1:** Calculate

\[
k = \ln\left(\frac{s_{m80}}{s_{m75}}\right)
\]

\[\text{………………… (2.9)}\]

This logarithm of the ratio of mortality rates is assumed to decline by a constant increment as age \( x \) rises above 80.

**Step 2:** Assign an arbitrary high value of \( s_{m75} + 0.66 \) to \( s_{m105} \).

In general, When LEB \( (e_0^0) \) is 70 years or higher (about 80 years) take \( \eta = 0.71 \)

When LEB \( (e_0^0) \) is 70 years or below (about 70 years) take \( \eta = 0.74 \)

**Step 3:** Estimate

\[
R = \frac{(6 * k) - \ln(\eta / s_{m75})}{15}
\]

\[\text{………………… (2.10)}\]

**Step 4:** Compute

\[
\begin{align*}
s_{m85} &= s_{m80} * \text{Exp}(k - R) \\
s_{m90} &= s_{m85} * \text{Exp}(k - 2R) \\
s_{m95} &= s_{m90} * \text{Exp}(k - 3R) \\
m_{100} &= s_{m95} * \text{Exp}(k - 4R)
\end{align*}
\]

\[\text{………………… (2.11)}\]

To test the reliability of the procedure of closing out mortality rates at older ages, Coale and Guo compared the rates calculated by above method with rates calculated by Gompertz method and with actual rates at ages over 80. They found that new estimates are closer than Gompertz estimates. Thus, the above method for closing out the life table to age 100 and above is giving good results than the Gompertz method. This method is recently used by the
Murray et. al. (2003). The “Modmatch” command available in STATA also used this method for predicting the older age mortality (Ferguson, 2002).

2.3.5 Method for construction of district level life table

It is observed that less number of deaths in small regions does not allow the development of life tables (Kouaouci et. al., 2005). Additionally, Congdon (2009) also stated that the application of life table method to small area data obstructed by problems such as small or zero death counts. Thus, one cannot expect to have a model life tables for smaller states and Union territories in India. To overcome this problem, a specific strategy is followed in this thesis, whereby the model life tables constructed using a group of bigger states of particular region can be assumed to provide acceptable district life table estimates of those states and union territories for which one cannot construct a model life table for the state under consideration.

As already stated that ASDRs are available only for seventeen major states of India; however, for remaining smaller states and Union Territories (which are Andaman and Nicobar, Arunachal Pradesh, Chandigarh, Chhattisgarh, Dadra & Nagar Haveli, Daman & Diu, Delhi, Goa, Jharkhand, Lakshadweep, Manipur, Meghalaya, Mizoram, Nagaland, Pondicherry, Sikkim, Tripura and Uttarakhand) ASDRs are not available. In addition, at district level only IMR (1q0) and U5MR (5q0) are available; ASDRs are not available. Therefore, for the construction of district level life table, thesis made one assumption that all the districts of a particular state are following the same mortality pattern as the state.
On the basis of above assumption, the thesis considers that the state specific life table will be applicable to the district of that particular state. Thus, to derive the district level life table, study enforces the regression models based on seventeen major states of India to the districts of those states.

However, as stated already, for smaller states and Union Territories thesis developed regression-based model life tables, based on the six groups which are classified on the basis of possible geographical contiguity, cultural and demographic characteristics, economic, health facilities and mainly the vital statistics (NIMS, ICMR and UNICEF, 2012). A description of the state group is given in Table 2.2.

**Table 2.2: Classification of the states on the basis of geographic, demographic and socio-economic characteristics**

<table>
<thead>
<tr>
<th>Groups</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG-1</td>
<td>Rajasthan, Chhattisgarh, Madhya Pradesh, Uttar Pradesh, Bihar, Jharkhand, Odisha and Assam</td>
</tr>
<tr>
<td>SG-2</td>
<td>Delhi, Haryana, Himachal Pradesh, Jammu &amp; Kashmir, Punjab and Uttaranchal</td>
</tr>
<tr>
<td>SG-3</td>
<td>West Bengal, Gujarat, Maharashtra</td>
</tr>
<tr>
<td>SG-4</td>
<td>Andhra Pradesh, Karnataka and Tamil Nadu</td>
</tr>
<tr>
<td>SG-5</td>
<td>Goa and Kerala</td>
</tr>
<tr>
<td>SG-6</td>
<td>Arunachal Pradesh, Meghalaya, Mizoram, Manipur, Nagaland, Sikkim, Tripura</td>
</tr>
</tbody>
</table>

*Source: NIMS, ICMR and UNICEF (2012).*

Since, for smaller states and union territories the ASDRs are not available, thus, to generate the life table for those states and henceforth districts, thesis adopted the technique of grouping of states. For instance group 1 (SG-1) is having eight states say Rajasthan, Chhattisgarh, Madhya Pradesh, Uttar Pradesh, Bihar, Jharkhand, Odisha and Assam. If one makes a group
by taking the average of ASDRs data, whatever data is available for all or some states, and develop a regression-based model and then the developed model may represent the average mortality pattern of any of the above states.

The group 6 (SG-6) is having all those states for which death rates are not available. So, to derive the district life table for those states (Arunachal Pradesh, Meghalaya, Mizoram, Manipur, Nagaland, Sikkim and Tripura) at the first step, the study considered the average of all major states. At the second step, it derives the regression-based model life tables (as developed for major states of India), and lastly employed these model life tables to the district of those states. In order to derive a model for SG-6, dissertation assumes the average of mortality pattern of all major states may represent mortality pattern of the smaller states. One may use even all India model earlier developed, however it is preferred here to take all the major states life tables as units (N=1242), just as in case of West model development of Coale and Demeny life table and also of United Nation’s General Model. In similar way, one can also construct the life table for those states as well as districts, for which death rate data by age are not available (Additional description is given in Chapter 6 of the Thesis).