Small Area Estimations

Need for developing periodic database and statistical procedures for obtaining reliable and valid estimates of health, demographic and other social sectors at the district and lower levels in India.

A Concept Note

K. Srinivasan

I. Introduction: Democratic decentralization and need for decentralized datasets and estimates of appropriate indicators.

An irreversible political process of democratic decentralization has been ushered in the country with the 73rd and 74th Constitutional Amendments passed by the Indian Parliament in 1992, forming the third tier of governance. Prior to 1992 the responsibilities of governance and allocation of needed resources, powers and responsibilities were constitutionally shared between the central and state governments. After 1992 a third tier of governance has been added through Panchayats in the rural areas and Nagar Palikas in the urban areas. Through the Panchayats and Nagar Palika Acts enacted after the constitutional amendments, these powers and responsibilities are, by constitution, shared with the local governments. Elected representatives at the village, block and district levels called the village Panchayats, Panchayat Samitis and Zilla Panchayats in the rural areas and Nagar Palikas in the urban areas have been constituted and functioning in most of the states in the country. Twenty-nine items of responsibilities that include primary education, basic health care, family planning, local water supply and sanitation have been allotted to these Pnchayats and Nagar Palikas. Allocation of needed resources for the effective implementation of these 29 items is however left to the state governments through the state finance commissions and supported by central funds, and the patterns of allotments vary from state to state, but the direction of democratic decentralization has been well defined and intensifying over the years. Program planning, implementation, monitoring and evaluation of many social development programs have to be necessarily vested in the coming years with the Panchayats and Nagar Palikas. The national and state level policies and district administration should have to play a facilitating and supportive role to these Panchayats and Nagar Palikas under the above scheme of things. The key governmental agency responsible for most of the social development programs, now rest with the district levels functionaries, elected for Panchayats and district level officials.

In this context there is a growing need, felt in many quarters of government administration and management of social development programs, to develop appropriate data bases, indicators and 'small area estimation procedures' for effective monitoring and evaluation of various social development programs including health, demographic changes, education, gender parity, employment and incomes at the district and lower levels. The need for development of appropriate databases at the levels, village, teshil/block and the district in the rural areas and the Nagar Palikas in the urban areas has become imminent. A district in India with an average population of about two million cannot be considered as a "Small Area" in the statistical sense and building data bases at this level seems mandatory for any objective planning, monitoring and evaluation of developmental programs. Estimation of various indicators at lower levels can be considered as falling in the domain of "Small Area Estimation" procedures. In this paper, I wish to highlight the major problems faced in the development of appropriate databases and indicators at the district and lower levels for the monitoring and evaluation of social development programs and provide a brief review of statistical procedures available in small area estimation of the selected indicators.

Databases: Current situation.

A. **Administrative Records (Service Statistics).**

In all the developed countries and in many developing countries, in recent years (e.g. Sri Lanka and China) the administrative records, called the "Service Statistics", compiled by the various departments of government, vital registration, health, education, family welfare, women's empowerment and employment, as an integral part of their supplies and services, form the major database for the routine monitoring and evaluation of various social development programs. Unfortunately in India such service statistics is woefully poor both in coverage and data quality.

One good illustration of the deficiency of service statistics is with regard to birth and death registration in the country. The Birth and Death registration Act was passed by the India parliament in 1969 making it compulsory for a birth to be registered within 14 days and a death within 7 days. The responsibility for the implementation of this Act rests with the state governments. Even after 50 years after the passage of the act it is rather distressing to note that among the births that occurred between 2000 and 2005, only 41 percent of the births were registered, 35 percent in the rural areas and 59 in the urban areas. Among the states the variation is very large, ranging from 85 percent in Goa to a low of 6 percent in Bihar, 7% in

Uttar Pradesh and 9 % in Jharkhand. Since the official responsibility for the registration of births and deaths is vested with the peripheral state government officials as the village Munsif/Patel, the ANM or male multipurpose worker, or the Panchayat secretary or the thane official depending upon the department of the state government with which the responsibility for the registration of vital events is vested, and since the vital events occur all over the place at all times, the extent of registration of births can be considered as a good proxy of the effectiveness of governance of the state government. It can be considered as a simple but good index of effectiveness of state governance (IEG). Viewed in this context, the IEG is highest, more than 80 % in Goa, Gujarat, Maharashtra, Himachal Pradesh, Kerala , Mizoram, Sikkim and Tamil Nadu and the lowest less than 20% in Rajasthan, Uttar Pradesh, Bihar and Jharkhand. None of the states in north (excepting Himachal Pradesh), central and east India, have good IEG, above 80 % registration.

Similar deficiencies in data sets compiled by the various ministries of the government, central as well as state, can be pointed out. Significantly, deficient in quality are the data with regard to enrolment of children in schools, health services offered and utilized, employment, identification of households below the poverty line and extent of utilization of various welfare services provided on subsidized basis to the population. It is sad and depressing to note that even after six decades of independence that the administrative records maintained by the various departments of health, education, employment and women's empowerment, and other social welfare departments cannot be used for routine monitoring and evaluation of programs and there is increased dependence on sample surveys.

B. Sample Surveys

Partly to offset the deficiencies in routine service statistics and to meet the imminent need for monitoring and evaluation of public programs of development, the scheme periodic sample surveys were adopted by the government and private or public institutions. For example the Sample Registration System (SRS) was initiated by the Registrar General of India in 1964-65 on a pilot basis and covered the whole country from 1968-69 in order to provide reliable estimates of birth and death rates and other measures of fertility and mortality including infant and child mortality and total fertility rates of the populations living in the rural urban areas for each state. Since 1969 these have become the major and probably only sources of reliable data for the routine monitoring and evaluation of fertility and mortality measures in the country and assessment of the impact of health and family planning programs. SRS is a dual recording system, wherein, in a random sample of villages and urban

areas selected from each state, all the households are covered by half yearly surveys to identify the births and deaths that occurred in these households during the past six months and simultaneously, a registrar of births and deaths is assigned for each village or rural block for continuous registration of the vital events. At the end of each year these two lists of vital events, those compiled from the half yearly surveys and the other from the continuous registration made by the registrar, are matched event by event and the unmatched vents are verified in the field for their validity with regard to location and timing of the vents. Then a full list of births and deaths that occurred in the SRS village or urban block is worked out. These data are used for computing the various vital rates at the state level for the rural and urban areas and published by RG.

When the SRS was started in 1964, it was clearly mentioned in the original document, that the SRS was a temporary scheme, to fill in the period of deficiencies in civil registration and should be in vogue for not more than ten years. Now, even after 40 years, the SRS continues to be the major source of data on births and death rates in the country and in 2004 the sample size has been considerably increased based on the 2001 census. While the SRS sample in the year 1969-70 covered a total of 3722 clusters, 2432 in the rural areas and 1290 in the urban areas, in 2004 the sample size was almost doubled to 7597 clusters, 4433 in the rural areas and 3164 in the urban areas. SRS seems to have got a permanent foothold in the monitoring and evaluation of births and deaths in the country and related health and welfare programs. The civil registration system seems to have suffered because of excessive focus on SRS and satisfaction of the researchers and bureaucrats with estimates of vital rates at the state level.

Similar observations can be made with regard to other surveys such as the National Sample Surveys, National Family Health Surveys and the District Level Household Surveys (DLHS) on reproductive and child health. Though these surveys have played an important role in providing basic information on health, nutrition, contraceptive use, utilization of health services and other parameters, they have in a way, instead of supplementing the various administrative records to be maintained and improved in the respective areas , have supplanted them becoming the major sources of data for program evaluation.

III. International monitoring of individual countries progress: Millennium Development Goals and India’s position.

IN the year 2000, 189 countries of the world, including India, signed an agreement to work towards achievement of specified goals called the Millennium Development Goals (MDG) aimed at global poverty alleviation, prevention of hunger, promote employment , provide basic education and health services and other sustainable developmental measures to improve the lot of the deprived segments of the population. There were eight goals, 21 related targets to be achieved by the year 2015 with the 1990 value as the base and 60 indicators to monitor the progress, these goals, targets and indicators were formulated by an Intergovernmental Committee of Experts constituted by the UN. The progress of the nations since 2000 is monitored by compiling data on all the 60 indicators and targets for each of the eight developmental goals from the countries. The latest MDG report of the UN provides data for the year 2008 and also provides a time series on the indicators from 1990 onwards up to 2008. We have thus a matrix of data for 60 indicators for 19 years.

These goals, and indicators are described in Table 1. These indicators offer a good framework for the development of minimal data bases and indicators for any country at different levels of its administrative set up. In India ,the Panchayats and administration from the district level and below have a crucial role to play in, planning ,implementation and the monitoring of the social development programs towards the first five goals, viz.;

Goal1: " Eradication of extreme poverty and hunger",
Goal 2: "Achievement of universal primary education",
Goal 3: " Promote gender equality and empower women" ;
Goal 4: " Reduce child mortality" and,
Goal 5: " Improve maternal health".

Towards these five goals, there are 8 program targets and 23 related indicators, that have to be monitored at the district and lower levels.

Table 1: Millennium Development Goals (MDGs)	
Goals and Targets (from the Millennium Declaration)	Indicators for monitoring progress
Goal 1: Eradicate Extreme Poverty And Hunger	
Target 1.A: Halve, between 1990 and 2015, the proportion of people whose income is less than one dollar a day	1.1 Proportion of population below $1 (PPP) per day[a] 1.2 Poverty gap ratio 1.3 Share of poorest quintile in national consumption
Target 1.B: Achieve full and productive employment and decent work for all, including women and young people	1.4 Growth rate of GDP per person employed 1.5 Employment-to-population ratio 1.6 Proportion of employed people living below $1 (PPP) per day 1.7 Proportion of own-account and contributing family workers in total employment
Target 1.C: Halve, between 1990 and 2015, the proportion of people who suffer from hunger	1.8 Prevalence of underweight children under-five years of age 1.9 Proportion of population below minimum level of dietary energy consumption
Goal 2: Achieve Universal Primary Education	
Target 2.A: Ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling	2.1 Net enrolment ratio in primary education 2.2 Proportion of pupils starting grade 1 who reach last grade of primary 2.3 Literacy rate of 15-24 year-olds, women and men
Goal 3: Promote Gender Equality And Empower Women	
Target 3.A: Eliminate gender disparity in primary and secondary education, preferably by 2005, and in all levels of education no later than 2015	3.1 Ratios of girls to boys in primary, secondary and tertiary education 3.2 Share of women in wage employment in the non-agricultural sector 3.3 Proportion of seats held by women in national parliament
Goal 4: Reduce Child Mortality	
Target 4.A: Reduce by two-thirds, between 1990 and 2015, the under-five mortality rate	4.1 Under-five mortality rate 4.2 Infant mortality rate 4.3 Proportion of 1 year-old children immunised against measles

Goal 5: Improve Maternal Health	
Target 5.A: Reduce by three quarters, between 1990 and 2015, the maternal mortality ratio	5.1 Maternal mortality ratio 5.2 Proportion of births attended by skilled health personnel
Target 5.B: Achieve, by 2015, universa access to reproductive health	5.3 Contraceptive prevalence rate 5.4 Adolescent birth rate 5.5 Antenatal care coverage (at least one visit and at leas four visits) 5.6 Unmet need for family planning
Goal 6: Combat HIV/AIDS, Malaria and other Diseases	
Target 6.A: Have halted by 2015 and begun reverse the spread of HIV/AIDS	6.1 HIV prevalence among population aged 15-24 years 6.2 Condom use at last high-risk sex 6.3 Proportion of population aged 15-24 years with comprehensive correct knowledge of HIV/AIDS 6.4 Ratio of school attendance of orphans to school attendance o non-orphans aged 10-14 years
Target 6.B: Achieve, by 2010, universal access to treatment for HIV/AIDS for all those who need it	6.5 Proportion of population with advanced HIV infection with access to antiretroviral drugs
Target 6.C: Have halted by 2015 and begun reverse the incidence of malaria and other major diseases	6.6 Incidence and death rates associated with malaria 6.7 Proportion of children under 5 sleeping under insecticide-treated bednets 6.8 Proportion of children under 5 with fever who are treated wi appropriate anti-malarial drugs 6.9 Incidence, prevalence and death rates associated with tuberculosis 6.10 Proportion of tuberculosis cases detected and cured under directly observed treatment short course
Goal 7: Ensure Environmental Sustainability	
Target 7.A: Integrate the principles of sustainable development into country policie and programmes and reverse the loss of environmental resources	7.1 Proportion of land area covered by forest 7.2 CO_2 emissions, total, per capita and per $1 GDP (PPP) 7.3 Consumption of ozone-depleting substances 7.4 Proportion of fish stocks within safe biological limits 7.5 Proportion of total water resources used
Target 7.B: Reduce biodiversity loss, achieving, by 2010, a significant reduction i the rate of loss	7.6 Proportion of terrestrial and marine areas protected 7.7 Proportion of species threatened with extinction
Target 7.C: Halve, by 2015, the proportion o people without sustainable access to safe drinking water and basic sanitation	7.8 Proportion of population using an improved drinking water source 7.9 Proportion of population using an improved sanitation facilit
Target 7.D: By 2020, to have achieved a significant improvement in the lives of at lea 100 million slum dwellers	7.10 Proportion of urban population living in slums

<table>
<tr><th colspan="2">Goal 8: Develop a Global Partnership for Development</th></tr>
<tr><td>Target 8.A: Develop further an open, rule-based, predictable, non-discriminatory trading and financial system

Includes a commitment to good governance, development and poverty reduction - both nationally and internationally

Target 8.B: Address the special needs of the least developed countries

Includes: tariff and quota free access for the least developed countries' exports; enhanced programme of debt relief for heavily indebted poor countries (HIPC) and cancellation of official bilateral debt; and more generous ODA for countries committed to poverty reduction

Target 8.C: Address the special needs of landlocked developing countries and small island developing States (through the Programme of Action for the Sustainable Development of Small Island Developing States and the outcome of the twenty-second special session of the General Assembly)

Target 8.D: Deal comprehensively with the debt problems of developing countries through national and international measures in order to make debt sustainable in the long term</td><td>Some of the indicators listed below are monitored separately for the least developed countries (LDCs), Africa, landlocked developing countries and small island developing States.

<u>Official development assistance (ODA)</u>
8.1 Net ODA, total and to the least developed countries, as percentage of OECD/DAC donors' gross national income
8.2 Proportion of total bilateral, sector-allocable ODA of OECD/DAC donors to basic social services (basic education, primary health care, nutrition, safe water and sanitation)

8.3 Proportion of bilateral official development assistance of OECD/DAC donors that is untied
8.4 ODA received in landlocked developing countries as a proportion of their gross national incomes
8.5 ODA received in small island developing States as a proportion of their gross national incomes

<u>Market access</u>
8.6 Proportion of total developed country imports (by value and excluding arms) from developing countries and least developed countries, admitted free of duty
8.7 Average tariffs imposed by developed countries on agricultural products and textiles and clothing from developing countries
8.8 Agricultural support estimate for OECD countries as a percentage of their gross domestic product
8.9 Proportion of ODA provided to help build trade capacity

<u>Debt sustainability</u>
8.10 Total number of countries that have reached their HIPC decision points and number that have reached their HIPC completion points (cumulative)
8.11 Debt relief committed under HIPC and MDRI Initiatives
8.12 Debt service as a percentage of exports of goods and services</td></tr>
<tr><td>Target 8.E: In cooperation with pharmaceutical companies, provide access to affordable essential drugs in develop. countries</td><td>8.13 Proportion of population with access to affordable essential drugs on a sustainable basis</td></tr>
<tr><td>Target 8.F: In cooperation with the private sector, make available the benefits of new technologies, especially information and communications</td><td>8.14 Telephone lines per 100 population
8.15 Cellular subscribers per 100 population
8.16 Internet users per 100 population</td></tr>
</table>

Table 2: Millennium Development Goals (MDGs) based indicators-India recommended to be computed at district and lower levels	
Goals and Targets (from the Millennium Declaration)	Indicators for monitoring progress
Goal 1: Eradicate Extreme Poverty and Hunger	
Target 1.A: Halve, between 1990 and 2015, the proportion of people whose income is less than one dollar a day	1.Proportion of population below $1 (PPP) per day (extreme poverty) 2.Percent below poverty line
Target 1.B: Achieve full and productive employment and decent work for all, including women and young people	3 Employment-to-population ratio 4 Proportion of labor force availing NREG benefits 5 Proportion of own-account and contributing family workers in total employment
Target 1.C: Halve, between 1990 and 2015, the proportion of people who suffer from hunger	6 Prevalence of underweight children under-five years of age 7 Proportion of population below minimum level of dietary energy consumption
Goal 2: Achieve Universal Primary Education	
Target 2.A: Ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling	9 Net enrolment ratio in primary education 10 Proportion of pupils starting grade 1 who reach last grade of primary 11 Literacy rate of 15-24 year-olds, women and men
Goal 3: Promote Gender Equality and Empower Women	
Target 3.A: Eliminate gender disparity in primary and secondary education, preferably by 2005, and in all levels of education no later than 2015	12 Ratios of girls to boys in primary, secondary and tertiary education 13 Share of women in wage employment in the non-agricultural sector 14 Proportion of seats held by women in national parliament
Goal 4: Reduce Child Mortality	
Target 4.A: Reduce by two-thirds, between 1990 and 2015, the under-five mortality rate	15.1 Under-five mortality rate 16 Infant mortality rate 17 Proportion of 1 year-old children receiving measles vaccine
Goal 5: Improve Maternal Health	
Target 5.A: Reduce by three quarters, between 1990 and 2015, the maternal mortality ratio	18 Maternal mortality ratio 19 Proportion of births attended by skilled health personnel
Target 5.B: Achieve, by 2015, universal access to reproductive health	20 Contraceptive prevalence rate 21 Adolescent birth rate 22 Antenatal care coverage (at least one visit and at least four visits) 23 Unmet need for family planning

Currently the data needs on health, reproductive health and gender related level at the state level are derived from the national level health and family welfare surveys, including the three rounds of National Family Health Surveys (1993-93, 98-99 and 2005-6 and the various rounds of National Sample Surveys. The district level household surveys conducted by the International Institute for Population Sciences, Mumbai provide data at the district levels from 1998-99 at four yearly intervals, mostly based on a sample of 1000 households for each district and cannot be used for estimation of relatively low prevalence variables nor for monitoring purposes on a more frequent interval for making program corrections on a periodical basis. The sample sizes needed for providing statistically reliable estimates of many crucial social development parameters at the local level become vary large; for example to get statistically reliable IMRs at the district level on a yearly basis we need at least 3 to 4000 households surveyed every year for births and deaths. The sample sizes will be substantially larger in the state of Kerala compared to Orissa and Bihar. Even the sample registration System that started with the idea of providing district level estimates at the district level, in their latest revision of samples in 2004 are now providing estimates only at the regional level (natural geographic regions covering many districts) within a state. If we have to estimate indicators such as the maternal mortality rate and prevalence of polio at the district level, which are rarer events then the sample sizes become very large. It appears that sample surveys cannot and should not supplant good service statistics but should only supplement them.

The three flag ship programs of the Government of India, the National Rural Health Mission (NRHM) implemented by the Ministry of Health and Family Welfare since 2005, the Sarva Siksha Abhyan (National Literacy Mission) and the National Rural Employment Guarantee scheme (NREGS) initiated in 2004 place a crucial participatory role for the people in the various public health programs, especially the maternal and child health program, and in employment guarantee for the unemployed labor force, through the elected representatives of Panchayats in the rural areas.

In this context it is appropriate and necessary that there is a planned and organized shift in the monitoring and evaluation of public health and other social welfare programs beginning first at the district level and then moving to the Panchayats in rural areas and Nagar Palikas in the urban areas. Village Panchayats cover small areas with 1500 to 2000 populations and at this level the only data that are currently available with some degree of reliability are the data from the primary census abstracts from the decennial population censuses. The birth and

death registration data, that should also be theoretically and legally available at these levels, are highly deficient in many states and of varying quality, with almost 100 percent registration of the vital events in Kerala and Tamil Nadu to less than 20% in Bihar and Uttar Pradesh.

There is an imminent need to build minimal data bases on education health, employment and related issues at the population and program levels beginning at these small population aggregates. The minimal data bases in the social development sectors are the following:

1) Births and deaths, as they occur on a continuous basis;
2) Number of births by parity of the mother and deaths by age and cause of death in broad categories;
3) Population size by age and sex, to be built annually, based on the latest census and subsequent births and deaths and estimated movements;
4) School attendance of children mainly in different age groups, separately for boys and girls;
5) Number of adult men and women employed, number taking advantage of employment guarantee schemes; and number seeking work, by broad age groups;
6) Number and details of those considered below the poverty line.

The data needed for the computation of the 23 indicators listed above can be accessed from the following sources.

1) The Sample Registration System
2) The National Sample Surveys.
3) Data from the service statistics of the education, health and family welfare departments
4) District Level Surveys on RCH, various rounds
5) Data from the National Sample Surveys
6) Vital Registration
7) NFHS-1,2 and 3rd rounds s
8) Data from the NREG

In order to compute the values on each of the 23 indicators listed earlier, the data sets from these sources have to be analyzed at the district level and below using the direct methods or various indirect methods available at that level. If the indicators based on the above data base are available only at a higher level of aggregation, such as the Panchayat Samiti or the district, and one is required to estimate the indicators at the Panchayat level, then one is required to use various procedures available under the indirect estimation techniques developed in demography or the various statistical methods developed under the “small area estimation” methods to derive estimates at the lower level. In addition to using the techniques

developed already, there is a growing need to adapt and modify the existing techniques to suit the data situations existing in India, in different states. The technique to be used should be consistent with the nature and quality of data available at that level for which estimation of the required indicator is needed and the data on the indicators available at higher levels of aggregation.

IV. Techniques of indirect estimation and small area estimations: a brief review.

The procedures used in Small Area Estimation Methods, can be broadly divided into two categories; demographic and statistical. The demographic methods can further be sub divided into three groups. They are 1) numerator or event analysis 2) regression analysis and 3) the standard indirect techniques of demography. The first method, numerator analysis is based only the numerator data on events with regard to their characteristics, in order to study the underlying process. For e.g, the parity or birth order distribution of births is a reflection of the population fertility levels, the distribution of infant deaths on the basis of their age at death is a reflection of the infant mortality levels. The underlying principle behind this approach is that every event has within its specified characteristics reflects the underlying demographic or social process. Further research work needs to be done in this area.

The second procedure viz. The regression approach is the most commonly used approach in demography to estimate an indicator, which is not known from a known set of other variables (called predictors) whose values are known. The regression is based on a number of assumptions using aggregated data sets at higher levels, data sets for comparable areas sets etc.

The third set of methods are commonly referred to as indirect methods and are those described in *Manual X* " Indirect Techniques of Demographic Estimation" brought out in 1983 by the Population Division Of the United Nations. This report is an aid to demographers and population experts to carry out the best possible evaluation and exploitation of data sources, especially those that are incomplete or deficient. It describes a wide range of time-tested techniques to make indirect estimates of demographic parameters. Each of the techniques presented is based on a mathematical model and explained in easy-to-follow examples. The computer program "Mortpak" also brought out by the UN, is in a way a companion volume to Manual X, giving a software package to derive indirect estimates of fertility and mortality from specified sets of data. Manual X describes nine methods as follows

I: Demographic models
II: Estimation of fertility based on information about children ever born
III: Estimation of child mortality from information on children ever born and children surviving
IV: Estimation of adult survivorship probabilities from information on orphan hood and widowhood
V: Estimation of adult mortality from information on the distribution of deaths by age
VI: Derivation of a smooth life table from a set of survivorship probabilities
VII: Fertility and mortality estimation using model stable age distributions
VIII: Estimation of fertility by reverse-survival methods and
IX: Estimation of adult mortality using successive census age distributions

All the above techniques are mathematical or algebraic and cannot be considered as statistical techniques. There are no requirements of sample sizes specified for the adoption of any of the above methods, nor did any sampling error estimate provided for any of the indirect estimates derive using any of the above methods. For example the methods of deriving estimates of fertility and child mortality from the data on average number of children ever born and children living at the time of the survey (average parity and average number of children surviving) classified by the age of the mother , commonly called as Brass methods (since developed by William Brass, a noted demographer) does not specify any minimum population sizes on which such averages should be based and theoretically can be applied to data from a few women from each age group. In general these methods are based on some apriori logic and algebra. For example in the Brass method of estimation fertility it is assumed that in any retrospective survey of women in a population, that while older women tend to underreport children ever born to them because of memory lapses or reluctance to report on child losses, they are not so with respect to number of children surviving to the time of survey. On the other hand they assume that the pattern of fertility reported in younger ages is more likely to be accurate. The ratio of reported average parity (P) and average cumulated fertility in the younger ages, up to age 25 (F) is used to correct the reported number of children ever born to get the completed family size. It is one correction factor, heuristically, derived. There are no sample size requirements or possibilities of estimating sampling errors in these

b) Statistical Procedures

There are two fundamental differences between the indirect technique of estimation of demographic parameters and the Small Area Estimation" methods developed during the past two decades by the statisticians. First the latter techniques are probabilistic with the estimates defined in statistical terms with variance and distributions. Secondly, more

importantly they are used with sample survey data where the sample sizes at the local levels for which estimates are required are small and mean values and variance based on them are not considered apriori statistically valid. The statistical theory and procedures in "Small Area Estimation" have seen a quantum leap during the past three decades and have been applied been applied effectively in the US and other countries in Europe in annual monitoring of poverty levels and changes at the county levels and changes in unemployment rates, that are relatively rare phenomena.

The recent book "Small Area Estimation" by (J N K Rao, Wiley) provides a succinct description of the various statistical methods available in this field. These techniques improve the estimation procedures by using models that establish some relationships between the areas, based on auxiliary information (census and/or administrative variables) related to the indicators variables of interest. These models provide in away indirect estimators that make use of related data from other areas which are comparable to the area under study and which might reduce drastically the estimation errors as long as model assumptions remain valid. There are four major statistical methods used in small area estimation and these are briefly described below. These are largely extracted from Wikepedia.

Generally the theory of resampling is used and includes a variety of methods. Common resampling techniques include bootstrapping, jackknifing and permutation tests. Estimating the precision of sample statistics (medians, variances, percentiles) by using subsets of available data is called **jackknifing** and drawing randomly with replacement from a set of data points is called **bootstrapping**. Exchanging labels on data points when performing significance tests is called **permutation tests**, also called exact tests, randomization tests, or re-randomization tests. Any models developed for validation of the estimates are also done using random subsets of the data (bootstrapping, cross validation). A brief description of these three methods are given below.

Bootstrap

"Bootstrapping is a statistical method for estimating the sampling distribution of an estimator by sampling with replacement from the original sample, most often with the purpose of deriving robust estimates of standard errors and confidence intervals of a population parameter like a mean, median, proportion, odds ratio, correlation coefficient or regression coefficient. It may also be used for constructing hypothesis tests. It is often used as a robust

alternative to inference based on parametric assumptions when those assumptions are in doubt, or where parametric inference is impossible or requires very complicated formulas for the calculation of standard errors”.

Jackknife

“Jackknifing, which is similar to bootstrapping, is used in statistical inferencing to estimate the bias and standard error in a statistic, when a random sample of observations is used to calculate it. The basic idea behind the jackknife estimator lies in systematically recomputing the statistic estimate leaving out one observation at a time from the sample set. From this new set of "observations" for the statistic an estimate for the bias can be calculated and an estimate for the variance of the statistic”.

Both methods (bootstrapping and jackknife) estimate the variability of a statistic from the variability of that statistic between subsamples, rather than from parametric assumptions. The jackknife is a less general technique than the bootstrap, and explores the sample variation differently. However the jackknife is easier to apply to complex sampling schemes, such as multi-stage sampling with varying sampling weights, than the bootstrap.The jackknife and bootstrap may in many situations yield similar results. But when used to estimate the standard error of a statistic, bootstrap gives slightly different results when repeated on the same data, whereas the jackknife gives exactly the same result each time (assuming the subsets to be removed are the same).

Cross-validation

“Cross-validation is a statistical method for validating a predictive model. Subsets of the data are held out, to be used as validating sets; a model is fit to the remaining data (a training set) and used to predict for the validation set. Averaging the quality of the predictions across the validation sets yields an overall measure of prediction accuracy.

One form of cross-validation leaves out a single observation at a time; this is similar to the jackknife. Another, K-fold cross-validation, splits the data into K subsets; each is held out in turn as the validation set.

This avoids "self-influence". For comparison, in regression analysis methods such as linear regression, each y value draws the regression line toward itself, making the predictions

appear more accurate than they really are, on average. Cross-validation applied to linear regression predicts the y value for each observation without using that observation.
This is often used for deciding how many predictor variables to use in regression. Without cross-validation, adding predictors always reduces the residual sum of squares (or possibly leaves it unchanged). In contrast, the cross-validated mean-square error will tend to decrease if valuable predictors are added, but increase if worthless predictors are added.

Permutation tests

"A **permutation test** (also called a randomization test, re-randomization test, or an exact test) is a type of statistical significance test in which a reference distribution is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points. In other words, the method by which treatments are allocated to subjects in an experimental design is mirrored in the analysis of that design. If the labels are exchangeable under the null hypothesis, then the resulting tests yield exact significance levels. Confidence intervals can then be derived from the tests. The theory has evolved from the works of R.A. Fisher and E.J.G. Pitman in the 1930s.Permutation tests are a subset of non-parametric statistics. The basic premise is to use only the assumption that it is possible that all of the treatment groups are equivalent, and that every member of them is the same before sampling began (i.e. the slot that they fill is not differentiable from other slots before the slots are filled). From this, one can calculate a statistic and then see to what extent this statistic is special by seeing how likely it would be if the treatment assignments had been jumbled.

For example the eradication of extreme poverty and hunger is the first of the Millennium Development Goals established by the United Nations. In order to plan for, monitor anti-poverty programs undertaken by the governments and evaluate these programs, it is necessary to have reliable statistical data on the living conditions of the people. However, sample sizes from many national surveys, including the National Sample Surveys and BPL surveys are not large enough to allow reliable estimation for small areas, such as the Panchayats and therefore small area estimation techniques that borrow information across areas through linking models are required. Linking models are based on auxiliary information such as census data and current administrative records. Many poverty measures are complex nonlinear functions of the values of a welfare variable for then. Then any of the techniques described above can be used.

Conclusions and Discussions

There is a growing need in India, to-day, to develop appropriate data base and methods of deriving necessary indicators for monitoring and evaluation of various social development programs at the district and lower levels, Panchayats and Nagar Palikas. The data provided by the Government of India annually in their monitoring of MDG goals, targets and indicators are grossly deficient on a number of indicators postulated by MDG.

Goal 1 to 5 of MDG and the 23 indicators related to them are a sufficient set of indicators for monitoring programs of poverty alleviation, primary education, primary health care and gender empowerment The data needed to compute these indicators at the district level are available from different sources and administrative records and surveys. Between the two primary attention should be given for strengthening of administrative records or service statistics, because only these data can provide necessary information for monitoring of programs at any level on a regular basis. The data from the surveys should primarily used for validation of the service statistics and more in depth causal analysis. But in India, for a variety of reasons, the major sources of data both for monitoring and evaluation of social development programs. the sample surveys have become the primary source of monitoring. This trend should be reversed by positive action to improve service statistics As an interim measure the available data from various surveys and statistics, mentioned in the earlier, should be utilized using both direct and indirect methods, demographic and statistical small area estimation procedures to get the estimates for each of the 23 indicators described earlier.

References

Good, P. (2005) Introduction to Statistics Through Resampling Methods and R/S-PLUS. Wiley. ISBN 0-471-71575-1

Hesterberg, T. C., D. S. Moore, S. Monaghan, A. Clipson, and R. Epstein (2005): Bootstrap Methods and Permutation Tests, software.

Battese, G. E., Harter, R. M. & Fuller, W. A. (1988) An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data, Journal of the American Statistical Association, 83, 28-36.

Elbers, C., Lanjouw, J. O. & Lanjouw, P. (2003) Micro-level estimation of poverty and inequality, Econometrica, 71, 355-364.

Foster, J., Greer, J. & Thorbecke, E. (1984) A class of decomposable poverty measures, Econometrica, 52, 761-766.

Gonz_alez-Manteiga, W., Lombard_a, M. J., Molina, I., Morales, D. & Santa-mara, L. (2008) Bootstrap Mean Squared Error of a Small-Area EBLUP, Journal of Statistical Computation and Simulation, 75, 443-462.

Rao, J. N. K. (2003) Small Area Estimation, Wiley, London.

United Nations, (1983) "Indirect Techniques of Demographic Estimation", Population Division