Potential Use and Pitfalls while using Survey and Administrative Data for Health Information

Dr. Udaya Shankar Mishra[1]

1. Professor, Department of Bio-Statistics and Epidemiology, International Institute for Population Sciences, Mumbai, India
   Email: umishra@iipsindia.ac.in

Table of contents

List of Tables:

Executive Summary

Comparison of survey statistics with administrative data often leads to disagreements raising concerns over the quality of survey on one hand and completeness of administrative data on the other Juxtaposing these two sources of data/information does not yield significant results owing to the exclusively distinct purpose they serve. Administrative data is denominator free information ideal for programme management and monitoring as opposed to survey statistics, although having similar information allows for a wider inspection. Administrative statistics can always be relied upon for trend inspection, coverage of intervention as well as programme monitoring and evaluation. However, if such an exercise is carried out without relying on the magnitude but rather on the indexed values with a given base, it may help avoid confusions and contradictions. On the contrary survey statistics has a wider significance. While survey statistics are often compared with alternative sources of data, it is essential to understand that surveys provide estimates and their validation is more in keeping with properties of these estimates based on features like confidence intervals, design effect and compositional effects. While there is indiscriminate comparison of survey- based estimates, for this paper we are taking into consideration instances when cross sectional estimates are compared with varying width of CIs and temporal comparisons are made overlooking compositional base of the outcomes. Hence when it comes to comparing administrative statistics and survey statistics, pattern matching is advisable. Further, great caution is required in analysing inter-survey estimates overlooking the aforementioned features of estimates. In conclusion, parallel reading of administrative statistics with survey estimates is not advisable unless there is a need to justify the accuracy and validity of such estimates. In such a case, they need to be bound by their purpose and evaluation features.

1. INTRODUCTION

Health statistics play a key role in decision making at all levels of health care systems. Extensive, timely, and reliable health and health-related statistics are fundamental for assessing the health sector and its transition. Health statistics provide evidence required for policy interventions and a better understanding of the health inequalities, social determinants of health, and epidemiological tendencies in the population. In addition, policy makers need to assess whether policies and programmes are directed towards the right beneficiaries, meeting set targets and whether appropriate monitoring and evaluation tools are in place. In this context, this paper attempts to understand the strengths and weaknesses of various official health statistics and health surveys available in India.

2. OBJECTIVES

The purpose of this paper is to reflect upon the usability of health statistics obtained from administrative data as well as from survey data in India. Specifically, this paper has the following objectives,

i. To list out the potential use of information obtained in surveys and administrative/service statistics.
ii. To list out the potential strengths of administrative/service statistics for monitoring and evaluation of programs in the absence of survey data.
iii. To describe the utility of surveys in understanding the dynamics of change along with its characteristics which remains limited in the case of service statistics.
iv. To evaluate the mutual robustness of both sources of information and compare the time-series feature of both.

3. SOURCES OF HEALTH INDICATORS

In a diverse country like India, it is extremely challenging to implement various programmes and policies. In the health sector, it is all the more strenuous to ensure that these programmes are in strict accordance with the policy guidelines and fall within the constitutional and legal framework so as to achieve the desired impact on all sections of the population, especially the underserved. To plan, design, implement and monitor these programs and to evaluate their performance and impact, statistical data on a variety of indicators is essential. Health indicators are summary measures that represent varied aspects of health or health system performance. Health indicators are generally developed as per a conceptual framework. In order to monitor the health system progress and performance, 47 indicators have been identified (WHO, 2011). Most countries select between 15 to 30 core indicators with baselines and targets to monitor their national health sector strategic plans. World Health Statistics 2019 focuses on the proposed health and health-related Sustainable Development Goals

(SDGs) and associated targets. It represents an effort to bring together available data on SDG health and health-related indicators.

Data collection for the health indicators must draw upon the full range of data sources. Each indicator needs to be linked to some data source in order to compile consistent estimates of the indicators. Data can be from surveys, administrative/secondary sources and primary sources. Some of the important health indicators in India along with their sources are listed in Table-1

Table 1: Sources of Important Health Indicators in India

Health Indicators	Source of Data
Fertility	Census of India, Sample Registration System & periodic surveys
Mortality	Census of India, Sample Registration System
Morbidity	National Sample Surveys & National Family Health Survey
Nutrition	District Level Household & Facility Survey, NFHS, CNNS
Programme Performance	National Sample Survey
Healthcare Infrastructure	Official Statistics Available in Health & Family Welfare Yearbooks

Some of the common data sources along with their coordinating agencies are provided in Table 2.

Table 2: Usability of Health Statistics in India

Source	Periodicity	Estimated Parameters	Levels of disaggregation	Usability
SRS	Annual, Since 1970	Fertility & Mortality Indicators like CBR, TFR, CDR, IMR, NNMR, PNNMR, U5MR, Sex Ratio (0-4)	State level estimates (big states), intra state regions and national level	Representative sample, Regular reports, Reliable.
Census	Decadal/10 Years	Population count by age, sex, area, IMR & Child Mortality	Population count down to village level, Mortality Rate at District level.	Reliable & valid population data in about 2 years. About 8 years lag for indirect estimation of Fertility & Mortality

CRS	Annual, 1958	Fertility & Mortality Indicators	District level and Sub-district level (large cities with 10000 population)	Yearly data for causes of death disaggregated by age, sex, type of residence; Data on trends related to vital events.
NFHS	6 Years I:92-93, II:98-99, III:05-06	Fertility & IMR (Indirect Estimates)	State level estimates	Indirect estimates. Quick estimates available within a year. Small sample size.
NFHS –IV	2015 – 2016	Fertility & IMR (Indirect Estimates)	District level estimates	Indirect estimates. Quick estimates available within a year. Increased sample size to around 6 lakhs to provide reliable estimates at the district level
National Nutritional Monitoring bureau	Annual	Nutritional Intake and Nutritional Status	10 States: Kerala, TN, Andhra, Karnataka, Gujarat, WB, MH, UP, MP and Odisha	
SAGE	DK	Health Problems and prevalence	For few states	
NSSO	Rounds	CBR, CDR, Health Problems and prevalence	National and State level	Provides reliable estimates.
Annual Health Survey (AHS) (Now merged in NFHS)	Annual	CBR, CDR, IMR, Neo-natal MR, U5MR, MMR, SRB, Sex Ratio (0-4), Sex Ratio (All Ages)	District level (284 Districts) of Only EAG States & Assam	To yield benchmarks of core vital and health indicators at the district level and to map changes therein on an annual basis.
Jansankhya Sthirata Kosh (National Population Stabilization Fund)	Annual	BPL Population in India. Literacy Rate in India.	State Level Health facility GIS Maps District Level Health Data	

WHO-Global Health Observatory Weekly epidemiological record. World health statistics quarterly. World health statistics Annual	Weekly Quarterly Annual	Wide variety of morbidity and mortality statistics	National level	

Source: Manual of Health Statistics in India, 2015

The sources of information on health and related issues have been grouped into two categories: indirect and direct sources. The indirect sources constitute Census, Civil Registration System, and Sample Registration System, while the direct sources encompass the sample surveys and official statistics, which are generally based on built-in information collected for any official programmes (Pandey, *et. al.* 2010). Population-based health surveys focus on service coverage, utilization, equity and population health outcomes. Population-based surveys are an invaluable source of health information. A prime purpose of these surveys is to provide high-quality data for policy development and programme planning, monitoring and evaluation. Population-based surveys have been used extensively to gather information on fertility, mortality, family planning, maternal and child health, and some other aspects of health, nutrition and health care in India (Dandona, R *et. al.* 2016).

4. HEALTH SURVEYS

Surveys differ from administrative data in terms of their purposes, and such differences often have implications for their statistical structure, conceptual framework, and content. A comprehensive list of major health surveys conducted in India during 1992 to 2016 and their sample size is given in Table 3. Almost all surveys are conducted to answer specific type of research or public policy questions as opposed to just fulfilling an administrative function. This difference in purpose is reflected in the population frame, the unit of observation, the sample size, and the scope and coverage of the data. In India major sources for health information obtained is surveys are[1]:

Table 3: Sample Size for Major Health Surveys in India, 1992 to 2016

---

[1] Garg, (2014) has well documented health surveys conducted in India by giving a succinct overview of various surveys, including the indicators they capture, the frequency, the focus and the like.

Survey	Survey years	No. of households in the sample
NFHS		
NFHS-1	1992–1993	88 562
NFHS-2	1998–1999	91 196
NFHS-3	2005–2006	109 041
NFHS-4	2015–2016	568 200
DLHS		
DLHS-1	1998–1999	529 817
DLHS-2	2002–2004	620 107
DLHS-3	2007–2008	720 320
DLHS-4	2012–2014	350 000
AHS		
AHS baseline	2010–2011	4 140 000
AHS 1st update	2011–2012	4 280 000
AHS 2nd update	2012–2013	4 320 000

Source: Dandona. R, et al., 2016

AHS: Annual Health Survey; DLHS: District Level Household Survey; NFHS: National Family Health Survey.

Census

Census is a decadal exercise conducted for monitoring population trends and features along with vital events like births, deaths and marriages during a one-year reference period. It is a complete enumeration of population about important demographic variables down to the district level. From 2001 census onwards important demographic variables began to be tabulated down to the panchayat level. Most of the information is aggregable by regions, gender and castes scheduled in the Constitution. Data is available at state and regional level and can be disaggregated by social categories. However, Census does not provide all the information required for assessing the health of the population such as morbidity, pattern of utilization of health services etc. (Pandey, et. al. 2010).

Sample Registration System (SRS)

The Sample Registration System (SRS) is a large-scale demographic survey for providing reliable annual estimates of birth rate, death rate and other fertility and mortality indicators at the national and sub-national levels by place of residence and gender. The Office of the Registrar General, India is the nodal agency in this regard. It is a continuous enumeration of births and deaths in selected sample units carried out by resident, part-time enumerators, generally *anganwadi* workers, teachers, independent survey investigators every six months, and by SRS supervisors. 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.

District Level Health and Facility Surveys

District Level Health and Facility Surveys - DLHS-1 (1998-99), DLHS-2 (2002-04), and DLHS-3 (2007-08) are household surveys conducted by International Institute for Population Sciences (IIPS), Mumbai under the guidance of Ministry of Health and Family Welfare (MoHFW). The DLHS-3 covered 611 districts in India, with 1000 to 1500 households chosen from each district. The focus of DLHS-3 was to provide health care and utilization indicators at the district level for the enhancement of the activities under National Rural Health Mission (NRHM). It estimated the coverage for antenatal care (ANC) and immunization services; extent of safe deliveries; contraceptive prevalence; unmet need for family planning; awareness about RTI/STI and HIV/AIDS; utilization of government health services and users' satisfaction. It also provides information on newborn care, post-natal care within 48 hours, role of ASHA in enhancing the reproductive and child health care and coverage of Janani Suraksha Yojana (JSY). An important component of DLHS-3 was the integration of Facility Survey of health institution (Sub centre, Primary Health Centre, Community Health Centre and District Hospital) accessible to the sampled villages.

National Family Health Survey (NFHS)

The National Family Health Survey (NFHS) is a large-scale, multi-round survey conducted in a representative sample of households throughout India. Five rounds of the survey have been conducted till now: NFHS-1 (1992-93), NFHS-2 (1998-99), NFHS-3 (2005-06), NFHS-4 (2015-16) and NFHS-5 (2018-19). NFHS-4 combined certain aspects of DLHS, by covering all 640 districts and 568,000 households. The National Family Health Survey is conducted by IIPS, Mumbai under the stewardship of MoHFW, with some financial and technical support from international agencies during the earlier surveys. This survey focuses on data on health and family welfare for policy and programme implementation. It provides state and national level information on fertility, infant and child mortality, the practice of family planning, maternal and child health, reproductive health, nutrition, anaemia, utilization and quality of health and family planning services. The NFHS-4 additionally covered perinatal mortality, adolescent reproductive health, high-risk sexual behaviour, safe injections, HIV, tuberculosis, and malaria, non-communicable diseases, and domestic violence. Each successive round of the NFHS has had two specific goals: a) to provide essential data on health and family welfare as required by the Ministry of Health and Family Welfare and other agencies for policy and programme formulation and implementation, and b) to provide information on important emerging health and family welfare issues.

Annual Health Survey (AHS)

The Annual Health Survey (AHS) is a three-year (2010, 2011 and 2012) periodic demographic survey conducted in nine high-focus states of Assam, Bihar, Jharkhand, Uttar Pradesh, Uttarakhand, Madhya Pradesh, Chhattisgarh, Odisha, and Rajasthan by Office of Registrar General of India. It is one of the largest population-based surveys in India. The second round of the AHS drew a representative sample of 20,694 primary sample units, covering 4.28 million households and 20.61 million people from 284 districts in these nine states. The objective was to set benchmarks of core vital and health indicators at the district level and to map changes therein on an annual basis. State level bulletins contain vital indicators viz. crude birth rate, crude death rate, infant mortality rate, neo-natal mortality rate, under five mortality rate, maternal mortality ratio, sex ratio at birth, 0-4 years and all ages. District level fact sheets contain 161 indicators on fertility, mother and childcare, family planning practices, mortality, disability, marriage etc. This data was released in 2011 and 2013 (ORGI 2013).

Coverage Evaluation Survey

Coverage Evaluation Survey is another population-based survey jointly carried out by MoHFW and UNICEF. Three rounds of this survey have been undertaken so far in 2005, 2006 and 2009. It covers all states and union territories and provides important information to assess the impact of NRHM strategies on coverage levels of maternal, new-born and child-health services. It specifically assesses routine immunization (RI); level of coverage of antenatal care (ANC), delivery care, and post-natal care (PNC); accessibility, availability and utilization of immunization and maternal care services including JSY; care-seeking behaviour for key childhood morbidities such as ARI and diarrhoea; initiation of breastfeeding, exclusive breast-feeding and complementary feeding; use of iodized salt; Vitamin A coverage; and the utilization of bed-nets by pregnant women and children.

Morbidity and Health Care Surveys

The Morbidity and Health Care surveys have been conducted in 1986-87, 1995-96 and 2004 by National Sample Survey Organization (NSSO), an arm of Central Statistical Organization (CSO) under the mentorship of Ministry of Statistics and Programme Implementation (MoSPI). The 71st round of the survey was completed by the NSSO in January 2014. NSSO conducts nationally representative, large-scale, multi-stage design survey to obtain detailed socio-economic data from the households on varied themes every year. Health care surveys are comprehensive household surveys that provide information on acute and chronic illness, hospitalization, expenditure on medicines and treatment, source of financing, socio economic status of individuals, demographic and educational profile of individuals, availability of drinking water and sanitation. The data is available by region, states, gender and socio-economic categories.

Consumer Expenditure Surveys

Consumer Expenditure Surveys (ConES) are carried out by NSSO, MoSPI quinquennially on a large sample of households. A small round with a smaller sample size is carried out every year with other NSSO socio economic surveys. The ConES is an important source of information on

affordability of health services including information on out-of-pocket expenditures and for analysis for impoverishment and catastrophic expenditures[2].

5. SCHEME OF SURVEY DATA

Some advantages of survey data over administrative data include the targeting of a specific population and variables of interest, the interaction with the respondent, and the ability to pledge that the data will be used solely for statistical (that is non-administrative) purposes. The potential challenges with survey data include difficulties in constructing a suitable frame, lack of legally mandated participation, high costs of increasing sample size, unit and item nonresponse and measurement error.

Researchers employing survey methods to collect data enjoy a number of benefits. Surveys are an excellent means to gather extensive information from households and individuals. Cost effectiveness of surveys depend on its extent of potential generalization and characteristic disaggregation. Since surveys enable researchers to collect data from a reasonably large sample with an optimum cost of execution, survey methods often adopt probability sampling techniques to generate estimates. To summarize, the advantages of survey data are (i) Cost-effectiveness, (ii) Generalizability, (iii) Reliability and (iv) replicability.

Similar to other methods of data collection, survey-based data also has its own share of limitation. Some may argue that surveys are flexible as they allow indefinite number questions on any number of topics. However, in reality, the survey researcher is generally stuck with a single instrument for collecting data (the questionnaire). Hence, surveys are in many ways inflexible.

Validity can also be a potential challenge in surveys. Survey questions are often standardized; thus, it may not be possible to diverge from the general questions. Consequently, survey results are often restricted to the methods and content of data collection and the uniform mean of exploration of this data adopted by the researchers.

While using survey data, it is rather important to identify the survey design to derive estimates with varying levels of disaggregation and characterization of outcomes. While levels of disaggregation depend on the sample size and the related outcome of interest, the characterization depends on the adopted stratification in the sample design. Further, comparability of the estimates derived based on survey data needs to consider the vital aspect of an estimate like the confidence interval, standard errors, differential sample size and above all characteristic composition.

---

[2] Van Doorslaer et. al. 2006, 2007; Garg and Karan 2009; Selvaraj and Karan, 2012 analysed NSSO Consumer Expenditure Surveys 1995-96, 2004-05 and 2009-10 for this purpose.

6. QUALITY OF SURVEY DATA

To fetch quality data from upcoming health studies and surveys such as National Family Health Survey (NFHS), the National Data Quality Forum (NDQF) a joint collaboration of ICMR's National Institute for Medical Statistics (ICMR - NIMS) and the Population Council was formed to identify gaps in data compilation and offer data quality solutions[3].

The NDQF attempted to identify issues in data quality. Lack of comparability and poor usability of national level data sources, discordance between system and survey level estimates, increased questionnaire length and questions on sensitive topics were the reasons that led to poor data quality. The NDQF also identified age-reporting errors or non-response and intentional skipping of questions, underreporting due to subjective question, interpretation and incompleteness and paucity of data to generate reliable estimates on mortality as major barriers to quality data.

According to ICMR, different public data sources report divergent numbers for the same indicator, for example, Sample Registration Survey (SRS) 2016 and NFHS conducted in 2015-16 report different sex ratios at birth and infant mortality rates, creating an interpretative dilemma. Incomplete information has been a challenge, for example, NSSO (2014) says that "Data that is generated at state level lacks any information on private sector where about 70% of population seek treatment."

With social silence on several issues, particularly sexuality, it is often difficult to elicit responses to sensitive questions in a large-scale survey in India. However, the NFHS-3 decided to incorporate several such questions, both for men and women (International Institute for Population Sciences and Macro International 2007). This made the NFHS-3 questionnaire significantly different, bulkier and perhaps more sensitive than those of the earlier surveys (Irudaya Rajan and James 2008).

The biggest shortcoming of health surveys is their inadequate number of poorly trained and poorly paid field agents to collect data who receive little logistical support and work under harsh conditions. This affects the quality of the data collected and the design of the survey itself.

It is widely accepted that the longer a questionnaire, the poorer the quality of data. In NFHS 4, the household questionnaire had 74 questions. The women's questionnaire was 93 pages long with 1,139 questions. Some of these questions were difficult for researchers to ask and for respondents to answer. The 40 questions related to domestic violence included questions on intimate sexual violence included questions like "how old were you the first time you were forced to have sexual intercourse or perform any other sexual acts by anyone, including (your/any) husband?" The men's questionnaire was 38 pages long with 843 questions. Of these 43 questions several sub-sections were dedicated to HIV and STD. NFHS 4 field researchers have repeatedly asked how it is possible to ask intimate sexual questions,

---

[3] See article, "*Why India lacks quality in its demographic and health data*?", The National Data Quality Forum Launch Event, Media Coverage Report, NDQF (2019), Available at https://www.ndqf.in/wp-content/uploads/2019/09/Media-Coverage-Report_29072019.pdf

especially when it is almost impossible to ensure privacy during the survey. The only answer they received was that questionnaire had been finalised and it has undergone pre-testing.

In their review of national health surveys, Dandona, *et. al.*, 2016 recommended a few modifications for the improvement of the design of the national health surveys. First, instead of having multiple, frequent surveys with overlapping goals, India should have a single major national health survey at five-year intervals. This could provide data on additional major causes of disease burden and their risk factors, along with cause-of-death data using automated verbal autopsy methods and include adult mortality rate estimation. The sample sizes should aim to provide state-level estimates for all indicators and district-level estimates for crucial indicators to capture the key features of health status heterogeneity across the country. Second, data collection on the key variables should be standardized to meet monitoring standards and to provide comparable data over time. Third, effective partnerships with a larger range of relevant stakeholders, including the academic community, should be established to increase the relevance and usefulness of the data. Fourth, detailed methods should be published. Fifth, individual-level data from these surveys should be made publicly available as soon as possible so that it can be used in the urgent tasks of informing policy and developing a more effective health system. Sixth, linking household survey data with health service use and administrative data, preferably using geospatial coding methods could be considered. Over time, India could also consider a continuous design for its national health survey, with advantages for survey management and timely provision of findings.

In the Occasional Paper published in the Observer Research Foundation (ORF) (Oommen, 2016) locates four major data gaps in the National Family Health Survey (NFHS)-4. First, there is a lack of data at the sub-state or the district level, making it difficult to plan for targeted interventions. Second, data is collected at irregular intervals, which does not allow for mid-course policy correction. For example, the NFHS-4 was conducted after a gap of nine years. Third, data remains incomplete in many surveys and tools, especially administrative data at hospitals and nursing centres in smaller towns and districts. Fourth, poor data quality at all levels. The issue briefly mentions that the lack of an independent quality control body limits the quality of available data, especially given that the information passes through various levels and processes before reaching the stage of evaluation and analysis.

7. ADMINISTRATIVE HEALTH STATISTICS

The Manual on Health Statistics in India (GOI, 2015) describes Administrative Data as information collected primarily for administrative (not for research) purposes. This type of data is collected by government departments and other organisations for registration, transaction and record keeping, usually during the delivery of a service. Administrative data is mainly collected for the following reasons -:

(a) Monitoring of government programmes and other forms of government intervention.
(b) Enabling regulatory activities and audit actions; and
(c) Targeting outcomes of government interventions.

Historically the main drivers for producing official statistics have been revenue raising, the support of policy formulation, and the delivery of public services. Industry and the private sector also rely on official statistics to inform their investment decisions. In its handbook on statistical organisation, the UN (2002) notes that although statistical offices "come in all sizes and many different shapes," they all process raw data, convert them into statistics, apply objective standards to their operations and make it a condition of survival to be impartial, neutral, and objective."

8. OBTAINING ADMINISTRATIVE DATA

Health information and statistics are important for planning, monitoring, and improvement of the health of population. However, the availability of health information in India is often in disarray. The following are the inadequacies in the current administrative data -

- The available information related to non-communicable diseases and injuries is poor. There is a significant gap as India is undergoing an epidemiological transition with these diseases /conditions accounting for a major share of disease burden.
- Information on infrastructure and human resources mostly comprises of the public sector overlooking the entire private sector that caters to disproportionate share of health services in India.
- Most of the information is disaggregated at the state level beyond which the information base is hardly compiled at the district level. This is undoubtedly a big limitation for the practical implementation of health programs at the district level under the proposed decentralization of health services in India.

Obtaining administrative records is often a challenging task. Data collection does not guarantee easy access to data. Administrative data may exist on paper and stored in various formats in different locations, however finding and digitizing these can be time consuming. Wherever digital records are available; data must be extracted (often by the only person in the organization trained to do so) before it can be shared and integrated into monitoring or evaluation system. This often takes a significant amount of time and must be planned for in advance. Turnover in staff who understand the data, know the data storage system, or who can make corrections to faulty data can complicate and delay these efforts. Gaining access to

usable administrative data in a timely manner may be difficult and requires will, coordination, planning, and resources.

9. QUALITY AND ISSUES IN USE OF ADMINISTRATIVE DATA

The use of administrative data depends on its process of collection and compilation. Although administrative data has been used in official statistics from the very beginning, its systematic collection and compilation for enhancing its potential remains neglected. Another important aspect is the lack of sensitivity of this source of information in terms of its utility beyond the concerned administration. The most important methods relevant to the use of administrative data are: (i) Data linking and (ii) Data modelling to overcome conceptual differences.

Data can be linked from different sources, across different levels, or over time. Linking data appropriately requires planning, preferably prior to data collection. Understanding linked data can provide depth and continuity to enrich otherwise discrete points of information. Although linking data is not necessarily expensive, there are nonetheless costs and benefits that should be considered while designing M&E plans and data collection activities. Why data linking? (i) Survey data sets (e.g., household and facility information) can be linked to compare services available and health outcomes across geographical units. (ii) Geographical and survey data can be linked to examine the effects of physical attributes on service utilization. (iii) Time series and panel data can help build causal explanations of program or project effects.

Conceptual modelling is an important activity for designing a database. The conceptual scheme is a concise description of data requirements specified by the application designer, including detailed descriptions about types of entities, relationships, and constraints (Elmasri R, and Navathe, 2011). Thus, the artefacts generated from the conceptual data modelling are important elements in building database systems. Currently, most Health Information Systems (HIS) are built using traditional database modelling technologies, in which both information and knowledge concepts are represented in single level computer systems using conventional data models (Marco E, *et. al.* 2005). However, HIS must handle a large number of concepts that often change or are specialized after a short period of time and, consequently, HISs based on such models are expensive to maintain.

One of the advantages of a decentralized statistical system is that the ministries and departments which are important contributors to statistical information are equipped with statistically qualified personnel. Consequently, there are minimal discrepancies in adopting and conforming to international standards and using statistical methods. The real issue in this context is not the understanding of the concepts and definitions, but the application of appropriate methodologies in collecting data in a highly heterogeneous country. This is one of the key responsibilities of the Central Statistical Organisation.

The quality of administrative data in both the health and education sectors have dropped alarmingly. In the past, hospitals and public health care system records and their counterparts in the public education system could provide fairly comprehensive information on these two sectors. However, with increasing private participation, especially amidst under-regulation the quality of coverage has dropped to such an extent that the data can no longer be relied

upon except perhaps as sentinel indicators. More recent public interventions, such as the Central Government's operation of the Integrated Child Development Scheme and the Government aided local schemes, have opened up new sources of administrative data, but their reliability is yet to be established.

There are two main challenges to the extensive use of administrative data. The first is that there is often a divergence between the nature of data required for administrative purposes, especially when the objective is to monitor programmes, and the nature of the data that would be required for statistical purposes. Since the data collection machinery is generally under administrative control of the programme authorities rather than the statistical authority, the introduction of appropriate questions and indicators quite often becomes a victim of the need to keep the record keeping process manageable. In response to this, the Government of India recently raised the status of the statistical officers in the line ministries significantly and, hopefully, over time their voice would be heard more prominently while designing the administrative data collection system. A similar effort is also underway to persuade the state governments to give more emphasis on statistical components of administrative records.

The second problem relates to the accuracy of the data. Although the completeness of coverage is frequently an issue, there is no significant problem in using the data if certain statistical corrections are made. Inaccuracy, however, can render the data completely useless for statistical purposes. By and large, it has been found that in situations where the data is collected for monitoring programmes, the quality of the data becomes highly questionable. This is a particular problem in the social sectors and in data that is collected by the taxation authorities. On the other hand, when the purpose is mainly for regulatory oversight the quality of data tends to be high.

The inadequacies of the existing Civil Registration System (CRS) are the incomplete registration of births and deaths especially in the rural areas, wide state-level variation in the registration of births and deaths and lack of data at district and lower levels especially on adult and maternal mortality. The main causes of poor registration are involvement of multiple line departments, lack of attention and priority for registration, lack of system for preparation and submission of statistical returns and lack of demand for birth and death certificate in schools and other places. Lack of awareness is also one of the reasons for low coverage of CRS (Gupta and Pandey, 2006).

Regarding privacy and confidentiality government agencies must ensure that personal information is not released publicly and is only available to authorised personnel on a need basis. It should be ensured that personal information cannot be derived from disseminated data and is maintained securely. Linking administrative data or allowing access to third parties opens up further layers of risk, including attacks on data systems, either from within organisations, data laboratories, or through the internet (if accessible in this way).

10. IMPROVING USABILITY OF ADMINISTRATIVE DATA

Given the vast size and diversity of India and the limitations and reach of the government machinery, administrative records will always tend to be incomplete here. The classic case of this is the coverage of the Civil Registration System, which is so low that it cannot be used for measuring demographic parameters between the census years. Many of these problems can be tackled through cross checks and corrections made through survey data. For instance, to overcome the incompleteness of civil registration, India operates a sample registration system which provides reasonable estimates of demographic indicators for the inter-census period. Similarly, surveys carried out by the National Sample Survey Organisation also provide important cross checks on a variety of statistical indicators.

Unfortunately, in a number of cases the survey data has supplanted the administrative data as the primary source of statistical information. Although the reasons are obvious, this is perhaps not desirable as the statistical agency supersedes the administrative operation while the two ideally need to be complementing each other. Further, survey exercises are much more expensive. A more efficient system would be to make less frequent use of surveys on strong statistical principles to provide validation and corrective factors for the data generated on a regular basis through administrative accounts. This would be akin to implementing a sample audit system, where perhaps the purpose would not be to find fault but to provide information which would be used to correct the inherent biases that may occur in administrative record keeping. In the final analysis, however, the main factors governing the usability of administrative data for statistical purposes are the legal framework underpinning the data collection activity and the political significance attached to the government interventions concerned and of course governance too.

11. HEALTH STATISTICS FOR MONITORING AND EVALUATION OF HEALTH PROGRAMMES

Monitoring and Evaluation (M&E) offers an understanding of the impact of an intervention, how well it is performing and whether it is achieving its aims and objectives; along with course correction if any need to be made for future intervention activities, and reporting on results used for both internal management purposes and for external accountability to fund agencies and stakeholders.

There has been a growing emphasis in India on Health Management Information System (HMIS) as a part of the National Rural Health Mission (NRHM) initiative to enable capturing of public health data from both public and private institutions in rural and urban areas across the country in order to strengthen the evidence-based planning of health programmes. Hence, one of the core strategies of NRHM in achieving its goals is to strengthen capacities for data collection, assessment and review for evidence-based planning, monitoring and supervision.

HMIS is designed to collect and report information on a programme, which allows managers at all levels to plan, monitor, and evaluate the operations and the performance of the whole programme. HMIS is a systematic process of collection, compilation, reporting, analysis and use of information on health care services. The information is generally helpful in planning, problem solving and decision making in health care service provisioning. Health management information incorporates all the data needed by policy makers, service providers/clinicians and health service users to improve and protect population health.

12. MONITORING AND EVALUATION: COMPARISONS OVERTIME

M&E Analysis uses data to make comparisons: comparisons over time (time series analysis), comparisons over space (cross-sectional analysis), and counterfactual comparisons (with/without project/programme). A key feature is to try and assess the extent to which government policies have had the desired impact on the problem that the original analysis identified, in other words to identify a causal link between policy and outcome which statistical analysis alone cannot achieve.

The most common use of the data obtained for M&E purposes is time series analysis. This involves the tracking of one or more indicators overtime to see the direction of change. The indicators themselves need not be complex but the prerequisite is a continuous supply of consistent and reliable data over the reference period. Good examples of suitable indicators are the service delivery indicators referred to earlier. The data may come from the service providers' own records, from focus group discussions and community surveys, or from random sample surveys of intended beneficiaries. Until quite recently national statistical offices were not all that involved in-service delivery monitoring and did not include the collection of service delivery indicators as a priority national information need.

Data or datasets that can be linked often reveal additional angles of interest for M&E, particularly for evaluation purposes. It is almost never possible to link data unless the linkage

has been planned in advance and built into the data procurement system. Time series and panel data are two ways of comparing surveys for monitoring and evaluation purpose. A time series is basically more than one set of data gathered in the same way from different samples drawn from the same population in a sequence of numerical data points in successive order. Panel data sometimes referred to as longitudinal data contains observations about the same cross sections across time.

13. HOUSEHOLD SURVEYS FOR MONITORING AND EVALUATION

The most popular instrument for monitoring is the household survey. It provides data that can be disaggregated to show results for different population groups and has the benefit of providing information both on the beneficiaries and the non-beneficiaries of development interventions. There are a number of different household survey models that can be used, each with its own strengths and limitations. The most complete coverage is provided by the decennial population census. Although obviously not appropriate for day-to-day monitoring, the census is vital because it provides the framework for almost all other household survey activities.

14. ADMINISTRATIVE DATA FOR PROGRAM MONITORING AND EVALUATION

Administrative data consists largely of financial and activity tracking data, which is the core data for many monitoring plans. Examples of these data include the number of chlorine dispensers distributed, number of trainings conducted, attendance rates at schools, or default rates on microloans. These can be used to monitor program performance, manage staff, and address implementation issues.

Administrative data may also include take-up, engagement, and feedback data. Utilizing these types of data can help managers and evaluators understand how individuals use and interact with the program. Administrative data often includes basic demographics or targeting data such as age, gender, marital status, business activity, or other demographic information that can be used to identify if the program is reaching its target population, or to identify possible differences across groups.

15. HEALTH MANAGEMENT INFORMATION SYSTEM (HMIS) FOR M&E

Health Management Information System (HMIS) is a Government to Government (G2G) web-based Monitoring Information System that has been put in place by Ministry of Health & Family Welfare (MoHFW), Government of India to monitor the National Health Mission and other health programmes and provide key inputs for policy formulation and appropriate programme interventions.

The main objective of Health Management Information System is to provide accurate, reliable and timely information to program managers and stakeholders for appropriate decision making. It acts as a tool for monitoring and evaluation of the program and based on the information available, appropriate planning can be done and executed for the people in need.

The government aims to build its HMIS as the main frame for integrating all other information and communication technology for health initiatives.

The government of India has been focusing on importance of HMIS and has emphasized on the quality of data so that the reports generated from the HMIS Portal can facilitate evidence-based decision-making process. States have taken various initiatives to improve the quality of data and among them one of the major initiatives is to conduct HMIS training (including Mother and Child Tracking Systems- E-Mamta) of Data manager, District M&E and block M&E assistant and Data assistant at CHC/PHC level on recording and reporting. In this context, State Health Society has developed training modules and reference materials for health workers. The main content of the same is HMIS formats, definition of data element, difference between recording and reporting register/ formats, MCTS reporting formats, use of data for Sub Centre level planning and technique of data validation. The reference manual for E-Mamta is printed in local dialect with user's friendly methodology.

16. DATA QUALITY ISSUES IN M&E

Data quality is a vital concern that needs recognition in M&E domain. To ensure the quality of the data collected, potential errors or biases in data collection or in the data itself, must be carefully considered in determining the usefulness of data sources and tools. Although problems in data quality usually require a technical solution, M&E plans must include discussions over data quality for any or all indicators where information or sources may be questionable. If the data on which the indicators are based are faulty, the indicators themselves cannot serve as a sound basis of program planning, management, monitoring, or evaluation. Any uncertainties about data sources or tools must be acknowledged and taken into account in the interpretation of all related results.

It is essential that the indicators used for M&E are 'fit for purpose'; that is, relevant to the needs of different users and sensitive to change. If health priorities, strategies or activities have changed, indicators should be reviewed to see if they are still relevant, and revisions should be made accordingly. The underlying data needs to be accurate, complete and timely. Quality is essential, both in terms of validity and reliability and finally transparency is critical.

17. ISSUES OF USING ADMINISTRATIVE DATA FOR M&E

There are three main issues relating to the use of administrative data sources. These are - limited scope, low data quality, and the challenges of obtaining the data.

Administrative data is usually available only for a specific group of people or clients. They typically exclude non-participants, those who were not eligible, or those who declined to participate. For instance, bank records will not have information on non-clients, which limits the types of evaluation questions that can be answered.

Administrative records are often restricted to a program's lifecycle, which may or may not be a problem. If administrative data is used to monitor inputs or if the outcome of interest is product take-up or usage, then administrative data may be sufficient. However, if the outcome of interest is expected to occur after the program ends—for example, if you are

interested in tracking employment outcomes for students of a job training program—this information would have to come from follow-up surveys of former students rather than administrative records. Often, administrative data comprises an important component of an M&E strategy but is rarely sufficient to fulfil all the strategy goals.

Poor data quality is a key challenge in using administrative data for M&E. The quality of data may be uneven and unknown, because data-tracking systems are often decentralized, and many have few quality-control mechanisms in place. Different units within the same organization may use different systems for collecting, checking, cleaning, and reporting information, resulting in inconsistency in quality across an organization. Sometimes those collecting or entering data falsify that data, particularly if they face incentives (such as rewards or bonuses) to meet targets. It is important to be aware of these challenges, investigate whether they apply to the relevant administrative data, and take steps to rectify the data quality problems they pose. If the data comes from within the organization, it may be possible to propose quality assurance measures and streamlining across departments as part of an overall M&E strategy. However, if data is obtained from an external source, it can be worthwhile to talk with them about data quality and can give some encouragement or instruction to improve.

The National Institution for Transforming India (NITI Aayog) which is the national body primarily responsible for implementing the SDGs in India has apprehensions about its ability to track and gather data for comprehensively evaluating the accomplishment of SDG targets. Lack of credible data will be a major roadblock for India in achieving the global goals. A strategy to address this concern could be a complete decentralization of the data collection process. The government could tap regional and local partnerships and build stakeholder capacities to gather and track data.

18. MEASUREMENT ISSUES OF SURVEY ESTIMATES AND OFFICIAL STATISTICS

Both administrative and survey data are used for decision making in health care services and outcome research. However, concerns exist about the quality and conceptual accuracy of administrative data. Additionally, there remains the potential challenge of under/overestimation of survey data due to various factors that may adversely affect policy making and research studies.

While comparing survey statistics and administrative statistics it should be kept in mind that we are comparing the incomparable with some aspects such as '*Confidence Interval*', '*Design Effect*' and the '*Composition Effect*'. As shown in the below Table-4, survey estimates have a confidence interval and have characteristics while official statistics lacks both. Survey data has characteristics like sex, age, religion, caste, etc. the composition of which may always have a bearing on any outcome when compared over time.

Table 4: Issues in Comparison of Survey Statistics and Administrative Statistics

	SURVEY STATISTICS	ADMINISTRATIVE STATISTICS
Confidence Interval		
Design Effect	Comparing the Non-comparable	Estimates without Confidence Interval
Composition Effect	Characteristic difference	Absence of characteristics

For example, comparisons of administrative and survey-based coverage estimate for child immunization programs indicate higher coverage reported from official coverage statistics. Comparisons of administrative coverage statistics and NFHS-4 survey data have revealed sizable discrepancies in coverage of child immunization. The NFHS-4 estimates of child immunization for BCG vaccination coverage given in Table-5 is 91.9 % for the period 2015-16 which is slightly higher than the official coverage statistics of 87% in 2015 and 88.7% in 2016. However, this difference in the two percentages could not be the basis of inferring that survey estimates are inaccurate.

Confidence Interval

For the estimates from survey data, sampling errors are used to construct confidence intervals, usually at the 95% level (or plus or minus two standard errors) to indicate a lower and an upper bound of the estimate. Irrespective of the methodology used, all surveys should provide estimates of full immunization coverage, which is comparable to the true population estimate of coverage. The NFHS surveys generate estimates at a given point of time and while trends in the status of child immunization can be statistically estimated from consecutive survey rounds, the official statistics, by advantage of its regularity of data recording, can provide relatively more accurate estimates of trends in immunization. However, calculation of the 'target population', i.e., all eligible children within a given area, which forms the denominator to calculate the proportion of children immunized, may have inaccuracies as it is estimated from the last census and data on new births available from government records. In addition, it is difficult to estimate in and out-migration while calculating 'target population'. This inherent difficulty in fixing a denominator when using official data is a known weakness in the system, which can be corrected to a large extent using statistical survey estimates and more accurate population estimates.

A confidence level is an expression of how confident a researcher can be of the estimates obtained from a sample. In practice, a confidence interval is used to express the probability that actual value

may lie between two given point for a parameter being estimated. There is uncertainty because inferences are based on a random sample of finite size from a population or process of interest. Standard errors can be used to work out the upper and lower limits (confidence interval) of the estimate, which will include the result from an equal coverage with a certain probability. The standard error is an indication of how close the sample survey estimate is to the result that would have been obtained from a census under the same operating conditions. Assuming that the target population is distributed normally for the characteristic being measured, the interval which contains the true value is usually calculated as being one, two, or three standard errors above and below the survey estimate. This interval is usually referred to as a confidence interval. If we want increased confidence, we have to take a wider interval.

Therefore, as per NFHS-4, as given in Table-5, we can say with 95% confidence that the child Immunization coverage for BCG vaccination in India is within 91.896% to 91.904% with a design effect of 1.644. However, this narrow range of the confidence interval practically is of no use for policy makers due to the very low value of the standard error giving an indication that the estimate is closer to the true population value.

Table 5: Child immunization coverage in India, NFHS-4 estimates and official statistics

Vaccination	NFHS-4 2015-16				Official coverage statistics 2015 & 2016	
	% covered (R)	DEFT	Confidence Interval		% covered 2015	% covered 2016
			R-2SE	R+2SE		
BCG	91.9 (0.002)	1.644	91.896	91.904	87.00	88.73
DPT 3 doses	78.4 (0.003)	1.52	78.394	78.406	87.40	88.45
Polio 3 doses	72.8 (0.003)	1.561	72.794	72.806	86.00	86.26
Measles	81.1 (0.003)	1.489	81.094	81.106	87.00	88.35

Note: Figures in parenthesis are Standard Error, R- Estimated Parameter, DEFT-Design Effect

Design Effect

In large sample surveys, in addition to the standard error, the design effect (DEFT) for each estimate is also computed, which is defined by Kish (1982) as the ratio between the standard error using the given sample design and the standard error that would result if a simple random sample had been used.

Design Effects (DEFT) compare the sample-to-sample variability from a given survey dataset with a hypothetical Simple Random Sampling design with the same number of individuals sampled from the population. DEFT is the ratio of two variance estimates. The design-based variance is in the numerator; the hypothetical Simple Random Sampling (SRS) variance is in the denominator. Therefore, DEFT is the ratio of two standard-error estimates. The design-based standard error is in the numerator; the hypothetical Simple Random Sampling with-replacement standard error is in the denominator.

The DEFT for a stratified sample is typically less than one, implying variance reduction due to stratification. In cluster samples, the DEFT is typically larger than one expressing the loss due to clustering. DEFT larger than one would imply an upward adjustment needed in the sample size calculated with a simple random sample. The sample size required for a cluster survey is almost always larger than that required for a random or stratified sample because of the DEFT factor.

Example of Design Effect

In a simple random sample of 50 households consisting of 120 persons, 27% were found to possess a mobile set. The sampling variances under a complex sampling design [$v(P_c)$] and simple random sampling [$v(P)$] of persons were computed to be 0.015 and 0.006, respectively.

The estimated design effect (DEFT) is

$$DEFT = \frac{v(P_c)}{v(P)} = \frac{0.015}{0.006} = 2.5$$

Hence the sample size must be raised to; n = 50 x 2.5 = 125. A sample of 125 households is expected to compensate for the loss of efficiency due to the use of the complex design. This loss arising out of using complex survey designs can sometimes be avoided by using what is known as the design effect (DEFT). It is defined as the ratio of the actual variance of the sample estimate obtained from a particular design to the variance of a simple random sample estimate of the same size.

Composition effect

There is a common tendency to read the statistics obtained in surveys over time as against the service statistics with disagreements between the two. Apart from such disagreements the survey estimates do not conform to a trend that is in keeping with interventions in play or the trends revealed by the service statistics. Such disagreements may not necessarily be due to reliability issues or data quality issues but issues of comparison of an outcome in the context of evolving transformation underway. In fact, inter-survey comparison needs to be adjusted for the composition of the base for which an outcome is measured. Such a composition could be sometimes vital basic parameters like age, duration or the like or it

may also be the characteristic composition which may not be the same between the two surveys. The need for a composition effect adjustment is essential for valid comparison given that we cannot ensure similar characteristic composition between two surveys for all possible outcomes. For instance, the debate on increasing stunting levels or anaemia levels may be an artefact of characteristic differences rather than the outcome per se.

19. DOS AND DON'T WHILE USING SURVEY DATA AND ADMINISTRATIVE DATA

The advantages and disadvantages of administrative data can be identified easily when they are compared with survey data. However, official data researchers are often face the challenge of mismatch between estimates obtained from various sources. The comparison of estimates from varying sources may not be worthwhile given their respective limitations and inadequacies. Administrative data is the information obtained primarily for administrative purposes. In most cases this type of data is collected by government departments or other types of organizations for the purposes of registration, transaction and record keeping, usually during the delivery of a service.

Survey data which is drawn from a sample and presented as statistics/estimates in principle should not be used for comparison with alternative statistics for a multiplicity of reasons. The first being an estimate should be valued subject to its features like confidence interval, the type of outcome (frequent or less frequent) and finally its characteristic composition. If the desired descriptive statistics is computed, it is possible to determine the stability of the obtained sample value. A sample represents the population from which it is drawn. It permits a high degree of accuracy due to a limited area of operations. Ultimately, the results of sampling studies turn out to be sufficiently accurate.

Of course, there are also a range of disadvantages of surveys that needs to be noted when using them. Surveys typically operate on a sample size approach where subsets of the overall population form the basis of enquiry. This means that you don't have data from everyone and generate some statistics to analyse the data effectively. Table 6 provides some basic guidelines to be kept mind on dos and don'ts while using survey estimates and official data.

Table-6 Guidelines on dos and don'ts while using survey estimates and official data.

Survey Data	Administrative/Official Data
• Confidence intervals locate the most likely range of the unknown population average.	• Such range of population parameter is not considered since there is no confidence interval for official data.
• Confidence intervals provide both the location and precision of a measure.	• Location and precision of measure does not matter/ is irrelevant while using official statistics.
• Smaller sample sizes generate wider intervals for the estimates to be considered less reliable. Hence, we compare the incomparable unless we account for the varying Cis.	• Administrative data represents the population and there is no concern about the sample size or interval width.

• The confidence interval is equal to two margins of errors and a margin of error is equal to about 2 standard errors (for 95% confidence).	• Margin of error is not applicable for official data.
• Design Effects (DEFT) compare the sample-to-sample variability from a given survey dataset with a hypothetical Simple Random Sampling design with the same number of individuals sampled from the population.	• Design Effects is not applicable for official data.
• Survey estimates are designed to minimise sampling and non-sampling errors.	• There are no sampling and non-sampling errors.
• Survey data has characteristics like sex, age, religion, caste, etc. which can be used to further classify the estimates according to these characteristic features.	• In most cases official data does not have characteristic features.
• Do not use survey data which does not have a suitable frame, lack of legally mandated participation, high costs of increasing sample size, unit and item, nonresponse, and measurement error.	• Data is collected by government departments or other types of organizations for the purposes of registration, transaction and record keeping for delivery of services and have certain legal mandate and lacks nonresponse, measurement errors, etc.
• Survey data is mostly cross-sectional data since it is mostly collected at a point of time.	• Administrative data has a time series as it is recorded at the time of service delivered throughout a year, or on monthly, weekly, daily basis.
• Difficulties in selecting a truly representative sample when the universe is too small or too heterogeneous.	• Not applicable.
• Validity can also be a problem with surveys. Survey questions are standardized; thus it can be difficult to ask anything other than very	• Such validity problems do not arise while using official data.

general questions that a broad range of people will understand.	
• Survey data does not pose any challenges in use if collected properly.	• Obtaining administrative records is often a challenge. Just because data have been collected does not mean it is ready to use.
• How is the sample defined? How large is the sample? Is the sample representative of the population of interest? What method was used for sampling?	• Such anomalies of sample size, representation of the population is not a concern of official data since it is recorded solely for administrative purpose.
• Survey data is more comprehensive and flexible than administrative data and is designed to capture a wide range of behaviours, preferences, and socio-economic indicators.	• This data rarely has detailed information on socioeconomic characteristics of the beneficiaries.
• Advantages of survey data over administrative data include the targeting of a specific population and variables of interest, the interaction with the respondent, and the ability to pledge that the data will be used solely for statistical (that is non-administrative) purposes.	• The advantages of administrative records as a data source are straightforward since those are consistently and accurately collected; resulting in highly reliable data covering a large number of observations, in some cases even 100 per cent of the population.
• There is systematic scheme of methodology, concepts and definitional clarity in survey data.	• Reviewing data collection methodologies is particularly important when using administrative data.

Dos and Don'ts of Survey estimate, and Estimate obtained from Administrative Statistics

	Survey Statistics		Administrative Statistics	
Domain	Dos	Don'ts	Dos	Don'ts
Reading Estimates	Read Estimates in consideration with CIs	Do not compare levels. Rather focus on comparing patterns	Read Temporal Trends	Do not compare with other sources in absolute terms but relate in an indexed manner.
	Limit characterization with concern for Design effect	Do not compare characteristic-based differences in estimates with other sources	Evaluate progress with temporal trends	Do not assess differences across regions in absolute terms
	Use it less for level comparison as against assessment of differentials and disparities	Temporal Comparison of Survey estimates need to be standardised by characteristic features	Use it in consonance with target population that will help in overcoming its denominator free feature.	Avoid comparison of levels rather than patterns across sources of similar statistics

20. SUMMING UP

Reliable data from health surveys are essential to describe the status and trends in health indicators by means of information not available from official registers. Population-based surveys are an invaluable source of health information. A key feature of these surveys is to provide high-quality data for policy development and programme planning, monitoring and evaluation. Meanwhile, administrative health data offers the possibility to study policy changes, intervention strategies adopted as well as differential coverage of varying interventions. As administrative data is not collected for the purposes of research, this data is generally messier and more complex than traditional social science datasets. Researchers should therefore not underestimate the amount of time and effort required to make it amenable for scientific analysis. Understanding the processes of how and why administrative data is collected will be central to assessing the data's quality and its suitability for social research.

Administrative data sets have their specific purpose for administering the programmes and interventions and need not necessarily be compared with Survey based estimates. Such comparison should be discouraged on two grounds i.e., the denominator free feature of administrative data and the estimation feature of survey data that compares only with consideration of the confidence interval, design effect and compositional effect. The denominator free feature of administrative data is also conditioned by the approximated denominator. While there is no scientific ground for comparison between the two sources of information in absolute terms, they may complement each other if patterns are read rather than the levels of outcomes. Further, temporal and cross-sectional comparison of statistics in the current SDG environment being routine, sufficient caution needs to be exercised in reading these statistics not beyond its intended purpose of generation. Ideally administrative statistics should be used for analysing patterns and survey statistics should be used for characterization. But beyond these designated purposes, comparisons should necessarily be guided by estimation properties for survey statistics and denominator approximation for administrative data. In conclusion, varied sources of information on the same phenomenon are not to contrast but to complement towards a better monitoring of programmes and interventions.

****

REFERENCES

Dandona Rakhi, Pandeya Anamika & Dandona Lalit (2016), "*A Review Of National Health Surveys In India*", Bull World Health Organ; 94:286–296A, https://www.who.int/bulletin/volumes/94/4/15-158493/en/

Elmasri R, Navathe S. B (2011), "*Fundamentals of Database Systems*", Addison-Wesley, 6th ed.

Garg C. Charu (2014), "*Availability, Quality and Data Gaps in Health Statistics in India*", Theme Paper, Ministry of Statistics and Programme Implementation (MoSPI), Government of India, http://www.mospi.nic.in/sites/default/files/reports_and_publication/technical_paper/Health_Statistics_17June2014.pdf

Garg Charu C and Karan Anup K (2009): Reducing out-of-pocket expenditures to reduce poverty: a disaggregated analysis at rural-urban and state level in India *Health Policy and Planning,* http://heapol.oxfordjournals.org/cgi/content/full/24/2/116

GOI (2015), "*Manual On Health Statistics In India,*" Central Statistical Office, Ministry of Statistics and Programme Implementation.

Gupta, R. K. and Pandey, A., (2006), Birth & Death registration status in Uttar Pradesh & Punjab. *Journal of Empirical Research in Social Sciences*, **1**(2): 43-45.

IIPS, Macro International (2007), "*National Family Health Survey India, 2005-06 NFHS-3*", International Institute for Population Sciences, Mumbai

Irudaya Rajan S and K S James (2008), "Third National Family Health Survey in India: Issues, Problems and Prospects", *Economic and Political Weekly*, Vol. 43, No. 48, November, pp 33-38

Kish, L. (1965), "*Survey Sampling*", New York, Wiley

Marco E, Thomas A, Jorg. R, Asuman D, Gokce L (2005), "A Survey and Analysis of Electronic Healthcare Record Standards"; *ACM Computing Surveys*, pp. 277–315.

National Sample Survey Organisation (2014), "*Concepts and Definitions Used in NSS*", Ministry of Statistics& Programme Implementation, Government of India.

NDQF (2019), "*Why India lacks quality in its demographic and health data?*", The National Data Quality Forum Launch Event, Media Coverage Report, Available at https://www.ndqf.in/wp-content/uploads/2019/09/Media-Coverage-Report_29072019.pdf

Office of Registrar General, India (ORGI) (2013), *Vital Statistics of India based on Civil Registration System*, 2009, Census, https://www.censusindia.gov.in

Oommen C. Kurian (2016), "*Overcoming Data Challenges in Tracking India's Health and Nutrition Targets",* ORF Occasional Paper 104, December.

Pandey Arvind, Roy Nandini, Bhawsar Rahul and Mishra R. M. (2010), "Health Information System in India: Issues of Data Availability and Quality", *Demography India,* Vol. 39, No. 1, pp. 111-128.

Selvaraj S and Karan AK (2012), "Why Publicly-Financed Health Insurance Schemes Are Ineffective in Providing Financial Risk Protection", *Economic and Political Weekly*, Vol - XLVII No. 11, March 17.

United Nations Statistics Division (2002), *Handbook of Statistical Organization, Third Edition, The Operation and Organization of a Statistical Agency, UN*, New York, http://siteresources.worldbank.org/INTPAME/Resources/Key-Readings/UN_manual.pdf.

Van Doorslaer E, O'Donnell O *et. al.* (2006), Effect of Payments for Health Care on Poverty Estimates in 11 Countries in Asia: An Analysis of Household Survey Data, *The Lancet* - Vol. 368, Issue 9544, 14, pp 1357-1364

Van Doorslaer E, O'Donnell O *et. al.* (2007), Catastrophic Payments for Health Care in Asia, *Health Economics,* Volume 16, Issue 11, pp 1159-1184

World Health Organization (2011), "Monitoring Evaluation and Review of National Health Strategies: A Country Led Platform for Information and Accountability".