

Where Is All Our Health Data Going?

SUNDEEP SAHAY, ARUNIMA MUKHERJEE

Multiple initiatives going on in India regarding the collection of digital personal health data are analysed, and the question of how the data is being used is examined. While such data could facilitate healthcare and referral services, a strong and sensitive governance structure is needed to be in place to enable its optimal use and to ensure that the data is not used to further the agendas of surveillance and control.

Contemporary times are characterised by discourses around “big data” and how these can be mined to extract value. The rapid proliferation of new information and communication technologies (ICTs) are making it possible to collect, store, combine, and share large volumes of data at almost real time levels of speed. We are at an early stage of this “data revolution” and as citizens whose personal data is at stake, it becomes important to understand how this data is collected, by whom, for what purpose, and its implications. As Zuboff’s (2014) digital declaration states:

When it comes to “big data” and the digital future, we are at the very beginning. Despite the rapid race of connection and the oceans of data it generates our societies have yet to determine how all this will be used, to what purpose, and who decides.

Zuboff (2018) calls data the “new oil” that is fundamentally transforming the economy and how companies interact with citizens, making them the means to further economic interests. While Zuboff’s critique focuses primarily on large western corporations like Amazon and Google, the role of data in public sector settings of the state in low- and middle-income countries (LMICs) like India is not analysed. Our analytical focus is on the use of personal data for public health settings in India. This increased focus on personal data is justified by the argument that it will help meet the challenges of poor health indicators, an endemic challenge. The question of interest then is what kind and volumes of health data are collected, and what remains its purpose. This article primarily focuses on the question of what data is collected, while arguing for the importance of understanding how it is used to improve conditions of health and well-being in India.

Increasing Health Data, But for What?

Modern health systems are increasingly dependent on good quality health data enabled through ICTs. Decisions on what data is collected, what information is generated from it, who has access to it, and for what purpose has huge implications for health outcomes and health rights. The choice of technology affects how data is collected, analysed, displayed, and to whom and how access is regulated. While ministries of health are primary users of data, there are increasingly major corporate interests involved in the development and management of data systems, for example, in the health insurance sector.

A health-rights perspective will view data systems as enabling the rights to access and verification of information by civil society, to enhance state accountability and to enable democratic decision-making and citizen engagement. At stake

Sundeep Sahay and Arunima Mukherjee (arunimam@gmail.com) are at the University of Oslo, Norway and Society for Health Information Systems Programmes, India.

are issues of privacy, confidentiality and data security, and how individual rights are challenged by the powers of an intrusive state or by market manipulations. Yet another major concern is how the use of data systems shapes the organisation of health services, and the relationships between providers and patients; this raises the question of whether the state continues to be the primary provider of health services to citizens, or acts only as a broker in the health marketplace.

The last decade has seen an explosive growth of the use of data and data systems in the Indian public health sector. But unlike sectors like banking and transport, the benefits of data systems are unclear, particularly to the citizens whose data it is. The initiative of strengthening health data systems as a key component of health sector reform was introduced in the 1990s, however, it is so beset with problems that it is in need of major reforms itself. This study gives an overview of data-related issues under the National Rural Health Mission (NRHM) starting 2008, and then traces current trends in gathering health data in both the public and non-public sectors. The aim is to understand from the perspective of citizens what data related to them is at stake, and how can they engage with it in safeguarding their interests and rights.

The Backdrop

While national health information systems (HIS) were introduced in India in the mid-1980s and 1990s, driven largely through the World Bank structural adjustment programmes, this article traces the story from 2008 soon after the NRHM was established in 2006. The NRHM's agenda was to bolster national public health systems including the health management information systems (HMIS) through architectural corrections of integration, rationalisation, and strengthening data use to improve health services delivery. The reform process was conceptualised in three stages.

(i) Phase 1 of situation assessment where HMIS data from three states were systematically examined showing, not unsurprisingly, a large amount of data (more than 3,000 elements) being gathered by each health facility, and less than 5% of it used to generate health indicators.

(ii) Phase 2 was of system redesign based on the application of the following practical design principles:

(a) No data should be collected more than once. For example, data on childhood tuberculosis was collected under HMIS, Integrated Disease Surveillance Programme (IDSP), and Expanded Programme for Immunization (EPI), with clear duplication and adverse impacts on data quality and its use.

(b) Only pertinent data that can generate relevant programme indicators should be collected. For example, prior to 2008, each data element was segregated by caste (Scheduled Castes [SC], Scheduled Tribes [ST] and others), used primarily to safeguard against potential questions in Parliament on issues of equity rather than to improve care.

(c) Data flows should be governed by a hierarchy of standards, where the lowest level (health clinics) retains the most granular data (for example, names and addresses of pregnant women), while higher levels should receive increasingly aggregated

data (for example, data to analyse the percentage of children immunised).

(d) Data reflecting something relatively stable (for example, population estimates of SC and ST groups) should be captured through annual surveys, rather than through routine monthly data, to reduce the data collection load.

(iii) Phase 3 was the implementation of the newly designed data collection formats (including 90% reduction in the data to be collected). The states received advice in the form of a letter from G C Chaturvedi, a visionary mission director, quoted below:

The Ministry of Health and Family Welfare has been taking steps to establish such a system for national level monitoring and also to help the States in building effective HMIS at the State, district and lower levels ... The department has developed a web based data entry portal (<http://nrhm-mis.nic.in>) which could provide analysis of data on *some specific indicators*, once the data is entered at the District level ... NHRM (National Health Systems Resource Centre) has an *open source software* also capable of handling local level information needs including data entry and analysis ... It is also available to States from NHRM **for free**. States willing to make use of it for Blocks/Facilities may avail the services of NHRM for implementing it. The HMIS formats to be used in this software are compatible with the Ministry's format. The software has the provision to *electronically* post the data from the Blocks/Facilities, compiled through it at the district level, directly to the national HMIS portal. (Chaturvedi 2009)

This letter touched upon various facets of the centralisation-decentralisation debate: (i) the centre should only receive data required to generate specific indicators (not all data collected); (ii) the centre should only receive aggregate district information; (iii) the use of free and open source software was encouraged; (iv) states should have the freedom to use software to collect subdistrict data; and (v) interoperability mechanisms are needed to enable state specific software to report data to the national database. This letter emphasised that health is primarily a state subject, and the centre should only request data to provide policy directions, while states should be free to use the software of their choice to report required data.

Unfortunately, this policy direction did not last long, for three key reasons. First, the central ministry positioned the national web portal as a "single window of truth" requiring all data to come there, ignoring the multiplicity of systems already existing in states. Second, this portal was designed to "prevent rather than promote interoperability" (for example, by not providing application programming interface [API] access), implying that state systems could not electronically share data with the national web portal. The message implied to states was that they could not use systems other than the national web portal. This lack of interoperability continues till today despite the advancements in technology, and government policy statements that mandate open software and access. The third reason was significant and path-breaking as Ghulam Nabi Azad, the then health minister, said he did not trust reported numbers on immunisation, and data quality would only improve when names were submitted. He is quoted as saying:

At present, the chief medical officer compiles the vaccine coverage data on the basis of what states tell him, without going into the field. So, we are not sure if the data is actually true. As a pilot project, we are

starting name-based vaccination with polio which is confined to just a few states. Once we have such data, we can simply call up a child's family and check whether he was actually administered the vaccine or not. (Sinha 2009)

This led to the birth of the megalomaniacal system, Mother and Child Tracking System (MCTS), which collected the names and details of every woman in the country registered for a pregnancy (about 22 crore annually) and similar numbers of children for immunisation. By design, the system planned to send all data collected to the national portal, despite only field health staff requiring detailed patient data to provide care. Clearly, it was a system based on the logic of strengthening surveillance (of the hapless health workers) rather than on improving care. This MCTS ran in parallel to the HMIS (contravening the principle of "no data should be collected more than once"), thus doubling data collection load and making health workers accountable for mismatches between the HMIS and MCTS data. As the entire focus of the government machinery shifted to promoting MCTS, the HMIS was significantly undermined. Data collected through the MCTS was outsourced to companies to call and follow up with the beneficiary or health workers to "authenticate" the data reported. There are various other issues regarding the MCTS, including those relating to cost benefits, ethics and legality of using personal data, and providing it to third-party vendors. These are larger topics to address that this study flags for further discussions.

Along with the MCTS, there was the explosion of other data collection systems. Various state-specific portals came up to meet the needs of states and donors. At national level also, new systems were introduced (such as the ASHA portal), significantly adding to the data load. Simultaneously, the central ministry moved relentlessly to get more decentralised data down to the individual level, in contravention to the 2008 circular requiring only district-level data. Today, about 3,00,000 facilities are reporting data monthly to the national web portal adding about one terabyte of data annually. The data collection load, primarily on the field-level health workers, is exponentially growing and is given minimal consideration in the design of new systems. S Sahay, P Nielsen, and M Latifov (2018) found that the auxiliary nurse midwives (ANMs) need to send (often overlapping) data to nearly 10 different systems and manage their data with 25+ primary registers (each about a metre long), resulting in increased administrative workload.

The Existing Situation, 10 Years On

Given this brief overview, it is pertinent to assess the volume of data and its use. It is, of course, determined by various factors (for example, governance, politics, infrastructure, capacity, etc [Rao 2017]), but is beyond the scope of discussion here.

Data load is constituted by aggregate facility-based and name- or case-based data systems. On an average, states report about 550 data elements through the national HMIS, and additionally have local data sets ranging from 2,500 to 650 data elements. Many data elements are inactive (no values reported). Inactive elements ranged from about 100 to even 2,500+ in different states. There are also programme-specific data

(mix of aggregate and case-based) for malaria, leprosy, blindness control, tuberculosis, and other diseases. Some states like Maharashtra, Odisha, and Uttar Pradesh (UP) have opted to build an integrated state data warehouse to store all their state-specific data to enable more integrated analysis. In most other states, these local data are collected in multiple systems, typically on paper or MS Excel files, with limited interoperability mechanisms. There are other data sets that this article does not analyse, such as for infrastructure, finance, human resources, and household surveys.

Case-based Data

Case-based data refers to individual-level data collected off a patient, including details of their demographics, programme-specific parameters, such as haemoglobin for ANC, and CD4 counts for HIV patients. The idea is for data to be captured longitudinally to improve point of care and strengthen referral linkages. This individual data is first compiled in field diaries, then transferred to primary registers, and then reports are compiled on paper, which are sent to the nearest computer location for data entry and reporting to the higher levels. In the past, monthly reports represented aggregates of the case-based data recorded in primary registers, but now, systems like MCTS require the entire patient record from the primary register to be entered in the computer and transmitted to state- and national-level servers. In most instances, the case-based data does not replace the aggregate reports (even though it is technically possible) and both continue in parallel, resulting in ambiguity on which data is the "truth." Very little feedback comes to the fieldworkers, who are left to wonder about the value and use of their reports. A summary of some programme-specific data is given in Table 1.

Table 1: Data Elements Collected under Various Programmes

Name of Programme	Total Data	Number	Total Data
	Fields	of Cases	Fields
	Per Record	Per Month	Per Month
National Programme for Prevention and Control of Cancers, Diabetes, Cardiovascular Diseases and Stroke	165	270	5,121
Maternal health	143	125	3,849
Eligible couple	91	95	980
Child health	61	250	4,437
Tuberculosis	59	12	227

Source: Compiled by the authors.

The existing systems are rampant with redundancies and also with important gaps. The health worker faces the brunt of reporting to multiple systems, with limited support and feedback. There is almost no data being made public to citizens.

How Is the Data Being Used?

Data is collected for different purposes, primarily to serve the bureaucratic function of monthly reporting to higher levels. Missing or "wrong" reports are subject to severe reprimand. More difficult to infer is how different functionaries use data to inform action to improve health services delivery and health outcomes. This can be inferred through proxies, such as: (i) the number of active data elements implying what they are reported upon; (ii) the number of data elements that are

systematically reported as zero or left blank; and (iii) the number of data elements used to generate indicators.

While the first proxy is indicative of elements currently relevant, the second one reflects if a health facility is actually providing a particular service. For example, if health facilities are systematically reporting the “number of low birth weight babies” as zero, then it could indicate that there are no low birth weight babies (which is highly unlikely), or that the weighing machines are unavailable, forcing staff to fill in “zeroes.” The third proxy reflects the principle of data for action, which assumes that if the data is converted to an indicator (in reference to a population denominator), it potentially becomes actionable. Whether this actionability potential is realised in practice is another matter, often subject to issues of political will, policy, governance, budgets, capacity, and infrastructure. Another challenge to actionability relates to our focus on numerator-only data, which creates challenges in “comparing to what?” Fredrick Mosteller (2010) illustrated this problem with a joke:

Why do the white horses eat more than the black horses?
Do not know. Why?

Because we have ten times as many white horses and black horses.

In public health data, denominators are crucial as all indicators are calculated with respect to a target population, such as the percentage of newborn babies immunised. In the absence of this denominator (“number of newborn babies for a period in a particular catchment area”), no meaningful intervention can be designed. In India, denominator figures are notoriously ambiguous, with the health and revenue departments having different figures and variations across health programmes and administrative levels. Private-sector data, which represents the bulk of healthcare services-provided in the country, is literally non-existent.

The data collection load for aggregate data (spanning national HMIS and state specific data) in a state varies between 3,000 and 8,000 elements per month. The number of inactive elements tend to be about 20%, while about 50% of the elements are reported as zero or blank on a systematic basis.

Table 2: Data Collected vs Data Used and Analysed

	Total HMIS Data Elements	Elements Report- ing Zeros	Percentage	Total State Specific Data Elements	State Specific Data Elements Reporting Zeros	Percentage	(Some) National Programmes Data Elements	National Programmes Data Elements Reporting Zeros
State A	550	273	50	2,536	2,400	95	580	490
State B	550	120	22	1,117	658	59	2,312	16
State C	550	300	55	661	530	80	–	–
State D	550	350	64	697	200	29	–	–

Source: HMIS portal.

Table 3: Total Data Fields for Case-based Data Required Every Month

Name of Programme	Total Data Fields Per Record	Number of Cases Per Month	Total Data Fields Per Month
National Programme for Prevention and Control of Cancers, Diabetes, Cardiovascular Diseases and Stroke	165	270	5,121
Maternal health	143	125	3,849
Eligible couple	91	95	980
Child health	61	250	4,437
Tuberculosis	59	12	227

Source: HMIS portal.

Furthermore, less than 10% of the data elements reported are used to generate indicators. Overall, this shows that we collect far more data than what is reported on, and an even smaller percentage is actually used for indicators (Table 2).

With respect to case-based data, Table 3 summarises all the data fields required every month from a peripheral provider (separate from HMIS aggregate data). There are exceptions such as programmes for malaria and leprosy that collect both aggregate and case-based data, as summarised in Table 4. They generate a healthy number of indicators, reflecting potentially an active use of data.

Table 4: Aggregate and Case-based Data Elements Collected

Programme	Aggregate Data Elements	Case-based Data Elements	Indicators
Malaria	660	178	134
Leprosy	29	168	55

Source: HMIS portal.

To summarise, the HMIS reflects a high degree of redundancy, with little systematic actual use of data, raising the question of “why is this data being collected?” Is this exercise only to satisfy bureaucratic requirements and to enhance the monitoring capacity of the ministry to “see like a state” (Scott 1997). But, the irony is that the state can drown in the sheer volume of data, especially when they receive all subdistrict data without abstraction. This, coupled with inadequate analytical and action taking capabilities at the centre, undermines even the bureaucratic function of reporting and contributes to data playing primarily a symbolic and legitimising function (Noir and Walsham 2007). Further, with data quality functions largely centralised, the principle of “data quality corrections are best done closest to the source of data collection” is violated. A large volume of data of suspect quality does not lend itself to effective use while magnifying the workload of field staff, who estimate that 60% of their time is spent on data related work.

These are some enduring data related problems:

- (i) The steady and huge increase in data collection seriously compromises the time health workers can spend on care provision.
- (ii) Despite this huge data collection, there is little reliable actionable information for use at the district level, with limited feedback to data providers other than reprimand.
- (iii) There is no provision for communities and public representatives to access data relevant to their participation.
- (iv) Introduction of case-based reporting has exacerbated problems of data overload and data entry, as this process adds and does not replace existing aggregate data flows. The workflow around case-based data is complex, often requiring the movement of cumbersome primary registers to distant offices where the computer is located.
- (v) Case-based data is collected without adequate privacy standards and data regulation frameworks, making it vulnerable to intrusive interventions, such as phone calls from call centres to households (or even their neighbours) to verify a pregnancy.
- (vi) There are multiple systems functioning with little to no interoperability, each working as an independent silo. Very recently, meta data and data standards (MDDS) were notified but the guidance on implementation is weak. Most applications, like the current HMIS, are expensive, proprietary and

rigid, and unable to change with changing needs and priorities leading to their early obsolescence.

(vii) The large investment in information technology (IT) in healthcare is not matched by outcomes. But despite this, there are little efforts at systematic evaluation and learning from past failures, and similar problems repeat in cycles.

(viii) Denominator data continues to be ambiguous or unavailable, leading to poor action taking potential of the numerator data being collected.

(ix) There is no data policy in place with regard to storage, length of time of storage, rights of access, and to ensure privacy.

(x) There is a repeated push for centralised electronic health records (EHRs), which is neither feasible nor desirable at a national scale. The standards being mandated (for example, Systematized Nomenclature of Medicine or SNOMED) are proprietary and Unites States-based, and have adverse implications for long-term and sustainable development of the architecture.

(xi) There is a push to develop gigantic systems at the national level, which will incorporate or render redundant all existing systems. Fortunately, while most do not take off, they tend to undermine existing systems in the process.

(xii) There are repeated pressures to create a single source of “truth” and outlaw other data and systems. While the accompanying rhetoric is to prevent redundancies, in practice this only adds to the problems and keeps data not transparent to the public.

(xiii) A fair number of population-based morbidity and mortality surveys are in use, which are welcome, but only some of the results are in the public domain.

Discerning Learnings from the NRHM Experience

The NRHM efforts to reform HMIS were only partially successful. A discussion on some broader design-related learnings extracted from this experience follows.

(i) Design systems to support the logic of care and not control: Monitoring systems are designed with the aim to control and discipline the performance of individual providers at the peripheral levels who, from fear of visibility, will tend to self-discipline anyway. The MCTS was a classic example of such a system.

(ii) Build governmental procurement systems that encourage the use of free and open source software systems: Governmental systems of procurement typically discriminate against free and open source software, and fail to develop contracts that enable the provision of dynamic and continuous support required to deal with the evolving informational needs. Procurement is influenced by large IT companies, whose marketing skills combined with a limited understanding of the health domain lead to inappropriate contracts and vendor lock-ins.

(iii) Build systems based on participatory design and a serious commitment to decentralise: Systems catering to what top managers think they need fail to factor in the requirements of providers who collect and enter the data to strengthen care processes. Quite often, IT reinforces bureaucratic control structures of power, and constrains rather than enables decentralisation.

While this discussion has focused primarily on data managed by the state, it is important also to understand data collected

by non-state actors like the private sector and international organisations. In India, this arguably is larger than public data. However, almost no published literature discusses this. In its absence, we use proxies to understand this data volume.

Health Data with Non-state Actors

Whether as a function of access, preference or economics, non-state actors play a critical role in the provision of health-care delivery in low- and middle-income countries. (McCoy and Briki 2010)

The collection and utilisation of health data is an integral aspect of the work of non-state actors, comprising national and international non-governmental organisations (NGOs), civil society organisations, philanthropic and commercial entities, individual medical practitioners, and pharmacies. They provide services in hospitals, nursing and maternity homes, clinics run by doctors, nurses, midwives and paramedical workers, diagnostic facilities, for example, laboratories and pharmacies (Mills et al 2002). There thus is great diversity in healthcare providers, including various non-state actors spanning non-residential and residential care, including hospitals, nursing homes, facilities for the elderly, rest homes for the mentally challenged, and those suffering from substance abuse (Kumar 2015).

In India, 50% of all hospital beds are in the private sector. The National Sample Survey Office (NSSO) of 2010–11 estimated about 10.4 lakh private healthcare enterprises to be active. This compares with 20% of private sector beds in Latin America, 30% in Asia and more than 34% in Africa (Hanson and Berman 1998). While this trend reflects efforts to optimise the use of non-state capacity, countries like Bangladesh have progressed slowly (MOHFW 2005), constrained by poor data availability. Building public–private partnerships has been an important strategy to expand healthcare provision capacity; for example, the Chiranjeevi Yojana in India demonstrated an increase in the institutional birth rate from 40.7% (2001) to 89.3% (2010), driven by sharp increases in private sector deliveries (Singh et al 2009).

In India, around independence, the public sector was the dominant provider of healthcare services, with the private healthcare sector accounting for only 8% of total patient care (Peters et al 2002). This situation has now dramatically changed, and it is estimated that 93% of all hospitals, 64% of beds, 80%–85% of doctors, and 80% of outpatient and 57% of inpatient services are catered to by the private sector (Tiwari and Nair 2006). The NSSO 2014 estimated that 72% rural and 79% urban spells of ailment were

Table 5: Contributions of Public and Private Healthcare Sector in Terms of Service Utilisation, Expenditure, Insurance Coverage and Infrastructure

Contributions	Public Sector		Private Sector	
	Inpatient	Outpatient	Inpatient	Outpatient
Services (%)				
1986–87	60	22.5	40	77.5
1995–96	43.5	19.5	56.5	80.5
2004–05	40	20.5	60	79.5
2014 (rural)	41.9	28.9	58.1	71.1
2014 (urban)	32	21.2	68	78.8
Contributions	Rural	Urban	Rural	Urban
Distribution of hospitalised cases (%)				
1995–96	44	43	56	57
2004–05	42	38	58	62
2014	42	32	58	68
Cost of per hospitalisation case (₹)				
1986–87	1,120	1,348	2,566	4,221
1995–96	3,307	3,490	5,091	6,234
2004–05	3,238	3,877	7,408	11,553
2014	5,636	7,670	21,726	32,375

Source: Sembiah et al (2018).

treated in the private sector. The National Family Health Survey (NFHS)-4 report estimated that 56.1% and 49% of members of surveyed urban and rural households respectively, sought private healthcare, while for the public sector, it was only 42% and 46.4% respectively (IIPS and ICF 2017). Table 5 (p 50) summarises these trends.

The estimation of data in the private healthcare sector can be made by the rate of services offered at the facilities. As per the 71st round of the NSSO (2014), the private sector accounted for two-thirds of inpatient and three-fourths of outpatient care treatments nationally. While private sector outpatient care has been nearly constant since 1986–87, inpatient care increased to 68% in urban areas and 58% in rural areas in 2014 from a low of 40% in 1986–87. Table 6 summarises these trends in percentage terms.

There is a variation in the distribution of private healthcare enterprises per 1,00,000 people across the states, with the highest numbers reported from the relatively high-income states like Delhi and Punjab, followed by states like West Bengal and Maharashtra that are higher than the national average. Table 7 summarises this statewide and sector-wide distribution.

Using this distribution as a proxy, it can be estimated that 60% of the total volume of health data is generated by non-state actors, which does not find visibility in the national HMIS (Bhattacharya et al 2012). This gap is of concern since the private for-profit sector represents more than two-thirds of human resources

Table 6: Inpatient and Outpatient Care Treatments by Type of Facilities

NSS Rounds	Description	Inpatient		Outpatient	
		Public	Private	Public	Private
42nd 1986–87	Total	60.0	40/0	22.5	77.5
52nd 1995–96	Total	43.5	56.6	19.5	80.5
60th 2004–05	Total	40.0	60.1	20.5	79.5
71st 2014	Rural	41.9	58.1	28.9	71.1
	Urban	32.0	68.0	21.2	78.8

Source: Various rounds of NSSO.

Table 7: Percentage of Distribution of Hospitalised Treatment across Sectors and States

State	% Persons Hospitalised			
	Private Hospital		Public Hospital	
	Rural	Urban	Rural	Urban
Andhra Pradesh	77.5	78.2	22.5	21.8
Assam	10.8	48.5	89.2	51.5
Bihar	57.4	61.2	42.6	38.8
Chhattisgarh	50.6	70.6	49.4	29.4
Gujarat	76.6	76.7	23.4	23.3
Haryana	66.7	81.7	33.3	18.3
Jharkhand	60.4	73.6	39.6	26.4
Karnataka	73.2	81.7	26.8	18.3
Kerala	65.3	66.7	34.7	33.3
Madhya Pradesh	46.5	58.3	53.5	41.7
Maharashtra	80.8	80	19.2	20
Odisha	18.7	42	81.3	58
Punjab	70.7	69.8	29.3	30.2
Rajasthan	45.8	45.6	54.2	54.4
Tamil Nadu	59.6	70.7	40.4	29.3
Telangana	71.4	78.8	28.6	21.2
Uttar Pradesh	69.8	71.7	30.2	28.3
West Bengal	22.8	47.4	77.2	52.6
All	58.1	68.0	41.9	32.0

Source: NSSO 71st round.

for health, including for maternal and child health services (Raban et al 2009), and in the absence of regulation or incentives there is no reporting taking place (Gautham et al 2016). It is arguably erroneously estimated that in 2014 about 75% of the total deliveries reported in national HMIS were from the public sector.¹

UP is the most populous state in India, with one of the highest maternal mortality rates (258 per 1,00,000 live births) and infant mortality rate (68 per 1,000 live births) in the country (2012–13 data; GOI 2014). The private sector in the state provides nearly 90% of treatment for acute illnesses, 80% for chronic conditions and 18% of institutional deliveries, representing 56.7% of all deliveries in the state. A health facility survey carried out during 2013 in 25 districts of the state reported that half of the 731 mapped private facilities providing institutional deliveries did not maintain any relevant records (KHPT and UoM 2013).

NGOs play an important role in India, and today, about 31 lakh are active. Multiple other international NGOs are active, supported by donors like the United States Agency for International Development. In one of the northern states, an NGO supported by an international donor is collecting health data through a team of about 600 staff, running almost a parallel health information system for the state, and their data is not part of the state repository.

Another rich source of health data is collected by medical colleges and research institutes who have designated study areas where they provide healthcare services and run the HIS, but often primarily for research purposes. The Indian Council for Medical Research (ICMR) has 31 institutes spread over the country researching different issues such as communicable diseases, fertility control, maternal and child health. Most of this data is collected in a project mode, and no central repository of this data exists. This is an important opportunity lost for data to be used for informing policy and studying disease trends over time. The ICMR is currently trying to strengthen their digitalisation activities and archive research data nationally. One such example relates to an antimicrobial resistance (AMR) surveillance research network, which includes a software tool to collect AMR data from 25 public and private hospitals and laboratories. There are around 40,000 records entered by two private lab chains in India in the last four to five months. An average volume of around 10,000 records are reported every month, which can become an active tool in the fight against AMR.

To summarise, huge amounts of health data are collected by the state, and even more by the non-state sector. State systems often focus on surveillance and control, not on care. Despite this huge data collection, there are important gaps, for example, AMR where no data is reported from 800 district hospitals for national surveillance purposes.

Collection of Case-based Data

The future promises an explosion in the collection of name/case-based data in India, with limited clarity on how this data would be used. Some planned initiatives are discussed.

(i) The Integrated Health Information Platform: Planned initially to replace the existing system for IDSP, it is now positioned as a platform to integrate all diseases, such as malaria and tuberculosis, and for one health data from all facility types. In short, Integrated Health Information Platform (IHIP) is positioned a system for everything and everyone, as stated in the website:

The primary objective of IHIP is to enable the creation of standards compliant EHRs of the citizens on a pan-India basis along with the integration and interoperability of the EHRs through a comprehensive Health Information Exchange (HIE) as part of this centralised accessible platform.

With the promise of improved continuity of care, confidential health records, and better diagnosis of diseases, the focus is very much on the individual rather than the population. The platform is being developed by two global private providers with expertise in human data science, banking, insurance and financial services, artificial intelligence, machine learning, cloud services, identity access, and security. The operationalisation of the platform is delayed, and in the meanwhile, efforts of departments like the National Vector Borne Disease Control Programme (NVBDCP) to develop a malaria surveillance system have been put on hold.

(ii) Health and Wellness Centres: The Ministry of Health and Family Welfare proposes to set up 1,50,000 centres as upgraded versions of existing health sub-centres and primary health centres to provide primary, outpatient, and referral care. The late former Finance Minister Arun Jaitley said,

will bring health care system closer to the homes of people ... for providing comprehensive health care, including for non-communicable diseases and maternal and child health services ... will also provide free essential drugs and diagnostic services. (Sundaraman 2018)

These centres are expected to handle more than 70% of all outpatient care. While nothing much has been written about the supporting HIS, they promise to focus on case-based data to capture the interaction between the patient and the local doctor. Another important issue to consider is the informational links between the HIS of these wellness centres with other national systems.

(iii) The National Health Protection Scheme: This scheme, also called Ayushman Bharat, was launched on 25 September 2018, and is described as being the world's largest insurance scheme, expected to cover approximately 500 million beneficiaries and 1.5 lakh health and wellness centres with coverage of ₹5 lakh a year per family. The scheme is expected to be cashless and portable across the country, and subsuming earlier schemes like the Rashtriya Swasthya Bima Yojana. The informational requirements to support such a scheme are immense and complex: (i) a functional EMR in every authorised health facility; (ii) a master-patient index with registration details of all citizens in one database; (iii) a family and individual identification scheme; (iv) interoperability between different systems; and (v) ability to include legacy data from earlier schemes with adequate security mechanisms. Many modern technologies are proposed, such as the cloud, big

data analytics, artificial intelligence, biometrics and augmented reality,² but their operationalisation will be complex.

(iv) The National Health Stack of the NITI Aayog: NITI Aayog (2018) published the National Health Stack (NHS) as

a visionary digital framework usable by centre and state across public and private sector ... that supports a multitude of health verticals and their disparate branches and is capable of integrating future IT solutions for a sector that is poised for rapid, disruptive changes and unforeseen twists. It is now conceivable to aim for digital health records for all citizens by the year 2022 ... Various layers of the National Health Stack will seamlessly link to support national health electronic registries, a coverage and claims platform, a federated personal health records framework, a national health analytics platform as well as other horizontal components.

The Stack's primary focus is on health insurance with plans to expand across other health programmes, promising analytics on average number of patients treated in a day, commonly occurring diseases, percentage of claims filed and fraud analysis, average billed amount by medical procedures, and average time of settlement.

In addition, there are various other initiatives on patient-level data. For example, states are implementing pilot models for universal health coverage and disease elimination programmes based on case data. The Centre for Development of Advanced Computing is developing personal health record systems, while the National Informatics Centre is rolling out hospital information systems. As citizens, we want to guard against the misuse of our personal data, and limiting its use to improve care services.

Conclusions

The levels of success starting from the use of HMIS for aggregate data have been limited, and many new complexities emerge for patient-level data. Citizens need to question where all the (individual)-level data is going, who will have access to it, and what they will be doing with it. Existing debates focus on the technology, and the topic of data use is largely marginalised. Some areas of citizen engagement include the following issues:

(i) Data systems should be developed on the guiding principle of reducing field staff's work burden whilst strengthening care processes.

(ii) Data systems must be based on free and open source software, where the source code remains with the end user with the required capacity. Contracts would allow space for dynamic development of the product with adequate data safeguards.

(iii) A policy on health data and its storage, management, and retrieval is required, along with provision for democratic access to data.

(iv) The EHRs should typically be maintained at the facility level, and only aggregated data stored in the cloud with supporting guidelines for anonymisation and access.

(v) The architecture must be decentralised but accompanied by measures that enable interoperability, incentives to encourage private sector involvement, and decentralised access to data.

(vi) Just as the individual has rights to access their personal data, communities (local, districts, states) would have

rights to access anonymised aggregate data pertaining to their area.

(vii) Data collected with public money from large surveys and used for policy purposes must be put up in the public domain for larger access to research.

(viii) Data collected through health and biomedical research institutions based on public money should be made available on the public domain.

NOTES

- 1 https://nrhm-mis.nic.in/hmisreports/frm-standard_reports.aspx.
- 2 <https://www.nhp.gov.in>.

REFERENCES

Bhattacharya, M, R Shahrawat and V Joon (2012): "Understanding Level of Maternal and Child Health Indicators Used in Health Management Information System among Peripheral Level Health Functionaries in Two Districts of India," *Journal of Health Informatics in Developing Countries*, Vol 6, pp 385–95.

Chaturvedi, G C (2009): "Letter to States," Do N-20020/2/2008-State (M&E), 9 November.

Gautham, M, N Spicer, M Subharwal, S Gupta, A Srivastava, S Bhattacharyya, B I Avan and J Schellenberg (2016): "District Decision-making for Health in Low-income Settings: A Qualitative Study in Uttar Pradesh, India, on Engaging the Private Health Sector in Sharing Health-Related Data," *Health Policy and Planning*, Vol 31, No 2, pp ii35–ii46.

GoI (2014): "Annual Health Survey Report: A Report on Core and Vital Health Indicators, Part 1," Office of the Registrar General & Census Commissioner, Ministry of Home Affairs, Government of India.

Hanson, K and P Berman (1998): "Private Health Care Provision in Developing Countries: A Preliminary Analysis of Levels and Composition," *Health Policy and Planning*, Vol 13, No 3, pp 195–211.

IIPS and ICF (2017): "National Family Health Survey (NFHS-4)," 2015–16, Mumbai: International Institute for Population Sciences, <http://www.rchips.org/NFHS/NFHS-4Reports/India.pdf>.

KHPT and UoM (2013): "Facility Mapping in 25 High Priority Districts of Uttar Pradesh: A Presentation Based on the Study," Karnataka Health Promotion Trust and University of Manitoba.

Kumar, S (2015): "Private Sector in Healthcare Delivery Market in India: Structure, Growth and Implications," <http://www.isid.org.in/pdf/WP185.pdf>.

McCoy, David and Nouria Briki (2010): "Taskforce on Innovative International Financing for Health Systems," *Bulletin of the World Health Organization*, Vol 88, pp 478–80.

Mills, A, R Brughla, K Hanson and B McPake (2002): "What Can Be Done about the Private Health Sector in Low-income Countries?" *Bulletin of the World Health Organization*, Vol 80, No 4, pp 325–30.

MoHFW (2005): "Chiranjeevi Yojana an Innovative Partnership with the Private Sector Obstetricians to Provide Skilled Care at Birth to the Poor in Gujarat," Ministry of Health and Family Welfare, <https://www.nhp.gov.in/sites/default/files/pdf/chiranjeevi-yojana-details.pdf>.

Mosteller, F (2010): *The Pleasures of Statistics: The Autobiography of Fredrick Mosteller*, Stephen E Fienberg, David C Hoaglin and Judith M Tanur (eds), Springer.

NITI Aayog (2018): "National Health Stack: Strategy and Approach," Government of India.

(ix) New initiatives must have clear plans on how they will engage with legacy data systems.

(x) Important data gaps, such as on AMR surveillance, need to be identified and plugged in.

Without the active engagement of citizens in these areas and more, the answer to the question "Where is all our health data going?" will be as Bob Dylan wrote, "The answer my friend is blowing in the wind."

Noir, C and G Walsham (2007): "The Great Legitimiser: ICT as Myth and Ceremony in the Indian Healthcare Sector," *Information Technology and People*, Vol 20, No 4, pp 313–33.

Peters, David H, A Yazbeck, G Ramana, R Sharma, L Pritchett and A Wagstaff (2002): "Better Health Systems for India's Poor: Findings, Analysis, and Options," World Bank Group, Washington, DC.

Raban, M, R Dandona and L Dandona (2009): "Essential Health Information Available for India in the Public Domain on the Internet," *BMC Public Health*, Vol 9.

Rao, S K (2017): *Do We Really Care? India's Health System*, New Delhi: Oxford University Press.

Sahay, S, P Nielsen and M Latifov (2018): "Grand Challenges of Public Health: How Can Health Information Systems Support Facing Them?" *Health Policy and Technology*, Vol 7, No 1, pp 81–87.

Scott, J (1997): *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*, Yale University Press.

Sembiyah, S, B Paul, A Dasgupta and L Bandyopadhyay (2018): "Capacity Building of Private Sector Workforce for Public Health Services in India: Scope and Challenges," *Indian Journal of Community Medicine*, Vol 43, No 3, pp 144–47.

Singh, A, D V Mavalankar, R Bhat, A Desai and S R Patel (2009): "Providing Skilled Birth Attendants and Emergency Obstetric Care to

the Poor through Partnership with Private Sector Obstetricians in Gujarat, India," *Bulletin of the World Health Organization*, Vol 87, pp 960–64, https://nrhm-mis.nic.in/hmisreports/frmstandard_reports.aspx.

Sinha, K (2009): "Azad Doubts Data on Child Vaccination," *Times of India*, 15 September, http://timesofindia.indiatimes.com/articleshow/5011257.cms?from=mdr&utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst.

Sundararaman, T (2018): "In New Health and Wellness Centres, India Has a Good Plan for Primary Care—Backed by Little Action," *Scroll.in*, <https://scroll.in/pulse/870405/in-new-health-and-wellness-centres-india-has-a-good-plan-for-primary-care-backed-by-little-action>.

Taskforce on Innovative International Financing for Health Systems (2009): *More Money for Health, and More Health for the Money*, World Health Organization.

Tiwari, V K and K S Nair (2006): "Private Health Care in India," *Health Millions*, Vol 32, pp 36–45.

Zuboff, S (2014): "A Digital Declaration," <http://www.faz.net/aktuell/feuilleton/debatten/digital-debate/shoshan-zuboff-on-big-data-as-surveillance-capitalism-13152525.html>.

— (2018): *The Age of Surveillance Capitalism: The Fight for the Human Future at the New Frontier of Power*, New York: Public Affairs.

EXPANSION

EPWRF India Time Series

(www.epwrfits.in)

Agriculture Census Statistics

Data sets from *Agriculture Census* have been added to the Agricultural Statistics module of the *EPWRF India Time Series (ITS)* online database. This sub-module contains state-wise data on:

- Number, Area and Average Size of Operational Holdings by Gender, by Social Groups and by Size Groups; and
- Characteristics of Operational Holdings by Tenancy Status, Terms of Leasing, Land Use, Irrigation Status, Sources of Irrigation and Cropping Pattern.

These characteristics are also provided in a two-way classification of Social Groups by Size Groups.

- Social Groups include Scheduled Castes, Scheduled Tribes, Others and Institutional Holders
- Size Groups are: Marginal (Below 1.00 hectare), Small (1.00 < 2.00 hectares), Semi-medium (2.00 < 4.00 hectares), Medium (4.00 < 10.00 hectares) and Large (10.00 hectares and above)

These data are available quinquennially from 1970–71.

Agricultural Statistics module constitutes one out of 20 modules of EPWRF ITS covering a range of macro-economic, financial sector and social sector indicators for India.

For more details, visit www.epwrfits.in or e-mail to: its@epwrf.in