

National rural drinking water monitoring: progress and challenges with India's IMIS database

James L. Wescoat Jr^a, Sarah Fletcher^b and Marianna Novellino^c

^a*Corresponding author. Department of Architecture, Massachusetts Institute of Technology, Room 10-390, 77 Massachusetts Ave, Cambridge, MA 02139, USA. E-mail: wescoat@mit.edu*

^b*Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, USA*

^c*Massachusetts Institute of Technology, Cambridge, MA, USA*

Abstract

National drinking water programs seek to address monitoring challenges that include self-reporting, data sampling, data consistency and quality, and sufficient frequency to assess the sustainability of water systems. India stands out for its comprehensive rural water database known as Integrated Management Information System (IMIS), which conducts annual monitoring of drinking water coverage, water quality, and related program components from the habitation level to the district, state, and national levels. The objective of this paper is to evaluate IMIS as a national rural water supply monitoring platform. This is important because IMIS is the official government database for rural water in India, and it is used to allocate resources and track the results of government policies. After putting India's IMIS database in an international context, the paper describes its detailed structure and content. It then illustrates the geographic patterns of water supply and water quality that IMIS can present, as well as data analysis issues that were identified. In particular, the fifth section of the paper identifies limitations on the use of state-level data for explanatory regression analysis. These limitations lead to recommendations for improving data analysis to support national rural water monitoring and evaluation, along with strategic approaches to data quality assurance, data access, and database functionality.

Keywords: IMIS database; India; Logistic regression; Monitoring; Rural drinking water

1. A perspective on national drinking water monitoring

Monitoring rural water coverage and quality at the national level poses challenges for all countries. Most wealthy countries, including the USA, have not produced comprehensive databases of local

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

doi: 10.2166/wp.2016.158

© 2016 The Authors

water system attributes and performance. India's Integrated Management Information System (IMIS), designed to monitor its National Rural Drinking Water Programme (NRDWP), constitutes an important exception to this pattern and is the focus of this paper.

National monitoring of rural access to drinking water has faced a number of systematic challenges since the 1977 Mar de la Plata Action Plan, which led to an emphasis on monitoring during the International Drinking Water Supply and Sanitation Decade from 1981 to 1990. Challenges include: (1) self-reporting of uneven and inconsistent data; (2) unsystematic sampling of water access, quantity, and quality; (3) inconsistent data metrics over space and time; and (4) simplistic distinctions between rural and urban. One expert went so far as to describe national data as 'nonsense statistics' (Satterthwaite, 2003).

Since that time, substantial progress has been made in improving national data monitoring. The United Nations Children's Fund (UNICEF) and the World Health Organization (WHO) have undertaken a Joint Monitoring Programme (JMP) that uses multiple samples and surveys. The JMP developed standardized survey instruments and common methods for reporting water and sanitation data discrepancies. Using these methods, India reported increases in rural access to safe drinking water from 64% in 1990 to 76.1% in 2000 and 90.7% in 2012 (WHO and UNICEF Joint Monitoring Programme, 2015). The JMP also developed common questionnaires and methods for compiling national and international datasets to estimate progress toward drinking water coverage goals (WHO and UNICEF, 2006). While the JMP constitutes a major advance over early self-reported percentages, it still relies on sample data and episodic (e.g. decennial) censuses, the specifications of which vary by source and time period. Interestingly, it does not appear that JMP data have incorporated national drinking water databases, such as India's IMIS.

At the national level, comprehensive efforts to monitor drinking water access are rare. The Rural Water Supply Network (RWSN), supported by the IRC International Water and Sanitation Centre, has a strong emphasis on monitoring (United Nations Economic Commission for Europe (UNECE), 2014; Schouten, 2015). RWSN notes that some countries in Africa have compiled national data (Ssozi & Danert, 2012), such as Ethiopia, but those datasets are not presently in the public domain (Sean Furey, pers. comm., 8 June 2015; United Nations Educational, Scientific and Cultural Organization [UNESCO]/WHO, 2015, p. 39). A new Water Point Database (Water Point Mapping, 2015) is being compiled in local areas of Tanzania on a voluntary basis. In addition, the IBNET water utility benchmarking database, initiated by the World Bank, provides information on a large and growing number of cities, particularly in Asia, Africa, and Latin America, but not for rural areas, or for all cities in a country (World Bank, 2015).

It is interesting to compare these international monitoring efforts in developing countries with those of wealthier countries such as the USA, which lags behind India on rural water supply monitoring. The US Geological Survey (Maupin *et al.*, 2014) estimates water use only by state and sector on a five-yearly basis. The US Census Bureau (2009) reports aggregate data on domestic water supply and plumbing systems of different types and sizes, but not their specific names and details, or detailed demographics of population served. The American Water Works Association publishes aggregate utility benchmarking data (Lafferty *et al.*, 2005), but no data for specific utilities (as IBNET does) due to utilities' confidentiality preferences. The US Department of Agriculture's Rural Utilities Service lists helpful support programs, and the National Drinking Water Clearinghouse compiles a large body of useful information online, but not on monitoring of specific rural water systems (Wescoat *et al.*, 2013). Where the USA does stand out in comparison with India and other countries is in its online Safe Drinking Water Information System, which reports on the water quality performance of drinking water suppliers (US Environmental Protection Agency (EPA), 2015).

An institutional challenge for national drinking water data collection is the constitutional primacy of states over water issues in federal systems of government such as Australia, Brazil, India, and the USA.

As an alternative, some state governments are creating their own drinking water databases, such as New South Wales in [Australia \(2014\)](#). Similarly in India, some states such as Maharashtra are supplementing IMIS with additional monitoring and evaluation data and tools ([World Bank, 2014](#)).

The aim of this paper is to assess the current capabilities and limitations of India's IMIS database. The next section reviews national-level research on drinking water in India, with an emphasis on uses of the IMIS database to date. The third section describes the IMIS database structure and methods used to assess it. The fourth section of the paper describes state-level national drinking water coverage and water quality patterns across states. The fifth section assesses the extent to which these national patterns can be explained through statistical analysis of IMIS state data, and a sub-state case study analysis in Gujarat, and it discusses the additional data needed to evaluate program outcomes. The concluding section of the paper identifies strategic priorities for enhancing national database development, analytics, and planning applications.

2. National drinking water monitoring and policy research in India

It is exceptional when a country invests in a full annual monitoring of drinking water supplies at the habitation level, as India has. This section discusses the scope and significance of this commitment, and reviews the evolution of India's drinking water programs and policies to date.

India's IMIS database stands out as an important example of a national drinking water monitoring system. A national drinking water database has many benefits because it:

- (i) documents all habitations, rather than a sample survey;
- (ii) provides descriptive data for policy planning at each level of government;
- (iii) offers insights into leading and lagging states, districts, and localities;
- (iv) sheds light on data gaps and quality;
- (v) enables statistical modeling for policy analysis.

IMIS water data are updated annually at the habitation level and aggregated at district, state, and national levels. The constitutional role of states in federal systems of governance may limit the scope and resources for national monitoring of local drinking water services in some countries, but India has managed to create a coordinated compilation of local, state, and national drinking water data. State and local organizations benefit by participating in national water monitoring because it is used for funding decisions. Consistent metrics enable comparisons of progress toward planning and policy goals, and sharing of experience and expertise on successful water and sanitation programs.

2.1. Evolution of drinking water programs in India

This section briefly reviews the development of India's national drinking water policies, which led to the IMIS monitoring database. Regulations for water and sanitation date back at least to the second century BCE, with the compilation of Kautilya's *Arthashastra*, or Book of Statecraft. It specifies the provision of water reservoirs for villages and animals; and the prohibition and fines related to pollution, poor drainage, and defecation near water bodies. This legacy continued in various traditions of customary law and practice that compile principles, proscriptions, and remedies for dealing with impurities in water and sanitation. However, the condition of water supplies deteriorated by the mid-19th century,

when colonial sanitation reformers in India and worldwide pushed for drinking water and hygiene standards, first for military cantonments and later for wider urban areas, supported by greater emphasis on collecting health and sanitation statistics (Harrison, 1994).

Upon Independence in 1947, Article 47 of the new Constitution of India asserted the duty of the state to improve public health and nutrition although it did not explicitly mention drinking water. Article 21 on the right to life has been interpreted as encompassing a right to water for basic needs, while the 73rd and 74th Amendments devolve responsibility and authority in principle to local governments. Water Aid (Khurana & Romit, 2009) has compiled a list of drinking water policies that we abridge and update in Table 1.

This survey of policies and related data sources indicate the significance of the shift to the IMIS national rural drinking water database in 2009, which moves beyond the reliance upon less frequent and less comprehensive data sources in earlier periods.

2.2. Literature search and review

This section of the paper reviews previous research on India's drinking water sector at the national level. State and local research is voluminous, but major national reviews that draw upon large datasets

Table 1. Drinking water policies in India.

Early Independence (1947–1969)

1949: Ministry of Health's Environment Hygiene Committee recommends provision of safe water for 90% of India's population in 40 years.

1969: National Rural Drinking Water Supply Programme is launched with UNICEF to provide bore wells, piped water supplies and related projects, following famine in Bihar.

Transition from technology to policy (1969–1989)

1972–73: Accelerated Rural Water Supply Programme (ARWSP) is created to increase the pace of state drinking water program funding and implementation.

1978: National water quality monitoring is begun by Central Pollution Control Board.

1986: National Drinking Water Mission (NDWM) is established under ARWSP following severe drought.

Restructuring phase (1989–1999)

1991: NDWM is renamed the Rajiv Gandhi National Drinking Water Mission (RGNDWM). 1991 Census provides drinking water data, followed by the National Sample Survey of 1993, and Demographic and Health Survey of 1993.

1994: The 73rd Constitutional Amendment assigns Panchayati Raj Institutions the responsibility of providing local rural drinking water.

1999: Department of Drinking Water Supply formed under Ministry of Rural Development.

1999: Total Sanitation Campaign is initiated to end open defecation. National Family Health Survey of 1999 provides data, as do the District Level Household and Facility Survey of 1999, and Multiple Indicator Cluster Survey of 2000.

Consolidation phase (2000 onwards)

2002: National water sector reform through the Swajaldhara program under the 10th five-year plan. 2001 Census provides water amenities data, as does National Sample Survey of 2002.

2004: Drinking water programs are brought under the umbrella of the RGNDWM. IMIS database is under development.

2005: Bharat Nirmal Programme created for rural sanitation and development.

2009: NRDWP begins, and includes the implementation of IMIS data collection.

2010: National Department of Drinking Water and Sanitation formed and becomes a Ministry in 2011.

2011–2022: Ministry of Drinking Water and Sanitation publishes a Strategic Plan for Rural Drinking Water.

2013: NRDWP Guidelines are updated (this is the current version).

2014: Drinking water and sanitation are encompassed in the Swachh Bharat Mission.

are few. A systematic bibliographic search was conducted using the search terms ‘India’ and ‘rural water’ in online indexes (WorldCat, Proquest Dissertations, Web of Science, Scopus, and Water Resources Abstracts) and grey literature sources (Government of India, UNICEF, Water Aid, and India Water Portal) (Wescoat, 2014).

The search indicated that early assessments used Census of India, National Sample Survey, and other periodic surveys in which drinking water is one of a large number of questionnaire topics. WHO and UNICEF (2006) prepared guidelines for local water and sanitation survey questionnaires. Local surveys usually do not have enough common variables for synthesis at the national level. At the regional scale, Prokopy (2005) collected local data on community participation and expenditures to compare two states’ water programs.

A transitional period occurred in studies that employed national data pre-dating the IMIS database. These studies estimated national drinking water coverage (Srikanth, 2009). Biswas & Mandal (2010) went beyond descriptive statistics to measures of correlation among drinking water variables. WaterAid (Khurana & Romit, 2009) compiled a historical perspective on rural drinking water policies and organizations in India, but relied upon Census data for descriptive statistics. The IRC developed a qualitative perspective on water supply service models and institutional analysis (James, 2011).

Although IMIS data became available from 2009 onwards, they have not been widely analyzed in national assessments. WaterAid’s (2011) ‘India Country Strategy 2011–2016’ includes propositions that could be tested through IMIS data analysis. Balasubramaniam *et al.* (2014) use econometric methods with Census data to draw inferences about the roles of caste and religion on differences in household drinking water access. Excellent reviews by UNICEF (2013) and Cronin *et al.* (2014) did not analyze IMIS data.

Studies that do draw upon IMIS data include a paper by Shrivastava (2013) on the presence of fluoride in drinking water. A national report by the Safe Water Network (2014) explores strategies for community water management, supported by IMIS as well as Census data. Cronin & Thompson (2014) discuss advances and limitations in the IMIS database, including data access, visualization, and quality. Most recently, Novellino (2015) examines IMIS in detail for rural water supply sustainability monitoring at the state and district levels, using Gujarat as a case study. As recommended by Cronin and Thompson, Novellino documents the data collection and compilation process, as well as data discrepancies, apparent data gaps, and detailed descriptive statistics relevant for analyzing slipback and sustainability. Here, we build upon Novellino’s research to show how IMIS data can be assessed in analytical and explanatory ways at the national scale.

3. Methodology and data

This section of the paper provides an analytical description of the IMIS database, based on a review of government documents, interviews with IMIS users and managers in Gandhinagar and New Delhi, and examination of online web content. The following section of the paper uses IMIS data to generate state-level maps and descriptive statistics for drinking water coverage and quality. The penultimate section of the paper then assesses the potential, and constraints, for using IMIS in explanatory statistical analyses to support policy and planning.

IMIS was launched in 2009 with the establishment of the NRDWP as a web-based platform to enable annual online monitoring of the status of water supply projects and coverage across rural India. IMIS

includes some historical data dating back to 2003. While historical records available within IMIS are limited, they will become a valuable resource for longitudinal analysis over time.

IMIS has four types of data for every habitation: habitation data (e.g. population, households, scheduled caste, scheduled tribe); scheme data (e.g. types of water storage, piped water supply, treatment, and costs); water source data (e.g. types of groundwater wells and surface water supplies); and water quality data (biological and chemical).

3.1. IMIS water supply and quality data

The habitation is a local community of households and is the smallest unit in IMIS. Habitations are classified as fully covered (FC), partially covered (PC), not covered (NC), and/or quality affected (QA). Coverage status is based upon the minimum national water supply standards of 40 liters per capita per day (lpcd) and 55 lpcd. The minimum quantity per person was 40 lpcd under the Swajaldhara water sector reforms program noted in [Table 1](#), and the next standard to be achieved by 2017 is 55 lpcd. The long-term goal for 2022 is to provide all rural areas with at least 70 lpcd of adequate water within the household or a 50-metre radius ([Department of Drinking Water Supply \(DDWS\), 2011](#)). An FC habitation has 100% of the population with adequate quantity and quality of water. If a habitation has quality problems, it is categorized as QA and therefore deemed NC regardless of the quantity of water available. A PC habitation must meet national water quality standards even if it has less than 100% of the population covered.

It is important to note that a habitation can have more than one water source and more than one water supply scheme. Thus, if a habitation is NOT categorized as FC, it means that ALL schemes for this habitation fail to meet the minimum requirements of water supply on quantity and quality. Similarly, if one water supply scheme fails, it does not necessarily mean that the habitation is NOT FC because there is often more than one scheme per habitation. If a water source fails, it means that any scheme entirely dependent upon this source fails. But if the scheme has multiple sources, then the scheme can remain functional.

3.2. Data entry and approval process

Data are entered at block, district, and state levels on an annual, monthly, or quarterly basis. The annual data entry is required for financial planning and budget allocation at central and state government levels. Annual data update the status of water coverage for all habitations in India (as FC, PC, QA, or NC). They also provide updated demographic data for habitations. After this survey is completed and annual plans are prepared, a group of projects is selected based on their priority and budget availability. These projects are called ‘Target Habitations and Schools’ and must be completed within the financial year.

Once the budget is allocated for annual target projects, monthly data are entered as progress reports (MPRs). The MPRs include infrastructure and financial data for ongoing and completed schemes, water quality of sources, community support activities, and operation and maintenance. Data entry is limited to district offices for the district MPRs. Based on MPRs, financial disbursements are approved and monitored at the state government level. The regular data entry process includes changes in sources, water quality facilities, and financial releases.

A small selection of users was interviewed to learn about the IMIS data entry process, and we found that they use IMIS as a required procedure for budgetary and accounting purposes. Few IMIS data entry officers download data for further analysis. Some keep duplicate data on separate spreadsheets at district offices. These duplicates may have formats that make it easier for district officials to keep track of their projects. Reasons for this practice include delays in updating the IMIS website, delayed website response, complex display of data on the website, lack of granularity of data below district level, and lack of familiarity with the full IMIS interface. This means that local users are not taking advantage of the full detail, functionality, and comparative power of the IMIS database.

3.3. Scope of the IMIS database

The discussion above is a simplified description of the IMIS database. The actual number of variables for each of the four main categories of IMIS is high (Novellino, 2015). The spatial scope of the IMIS database includes all geographical divisions in India: national, state, district, block, panchayat, village, and habitation. Field surveys performed at the habitation level are aggregated to create district-level data. Data for some formats are not collected at all spatial levels, resulting in limitations on local data analysis. Even when habitation data are available, it is only by drilling down through district and block tables. Compiling data across larger administrative areas entails downloading and reassembling myriad habitation-level tables, a major limitation for national program evaluation and policy analysis. Ready access to local data across administrative areas is limited to the central government.

4. Major national observations using the IMIS database

With this understanding of the IMIS database structure, we now use it to analyze patterns of drinking water coverage, investment, and water QA habitations.

Figure 1 compares investment with coverage, both in terms of expenditures and schemes built. Cumulative public expenditures increase from FY 2010–2011 to FY 2014–2015. However, while annual state expenditures remain steady, annual national expenditures decline slightly in absolute and percentage terms, which may reflect a trend in financial devolution. Spending is correlated with the number of schemes built (middle line), but curiously those expenditures and schemes do not appear to have had a significant impact on the percentage of habitations that are fully covered, particularly at the 40 lpcd level.

To dig deeper into this trend, we look at expenditures by state from April 2010 to March 2015. We find that the proportion of state and national funding varies considerably across states, with Sikkim, Punjab, and Nagaland receiving more than 95% of their expenditures from the national government while some states, notably Gujarat, provide more than half of their own expenditures. Expenditures are highest in Rajasthan and Karnataka (12% and 10% of total national expenditures, respectively), which have arid or semi-arid conditions with regular water shortages, followed by the relatively large states of Uttar Pradesh, Gujarat, and Maharashtra (9%, 9%, and 8% of total national expenditures, respectively).

4.1. Current coverage status

Figure 2(a) displays the percent coverage at the 40 lpcd level for each state and union territory, broken into deciles. Figure 2(b) shows the breakdown of habitations into fully covered, PC, and QA categories

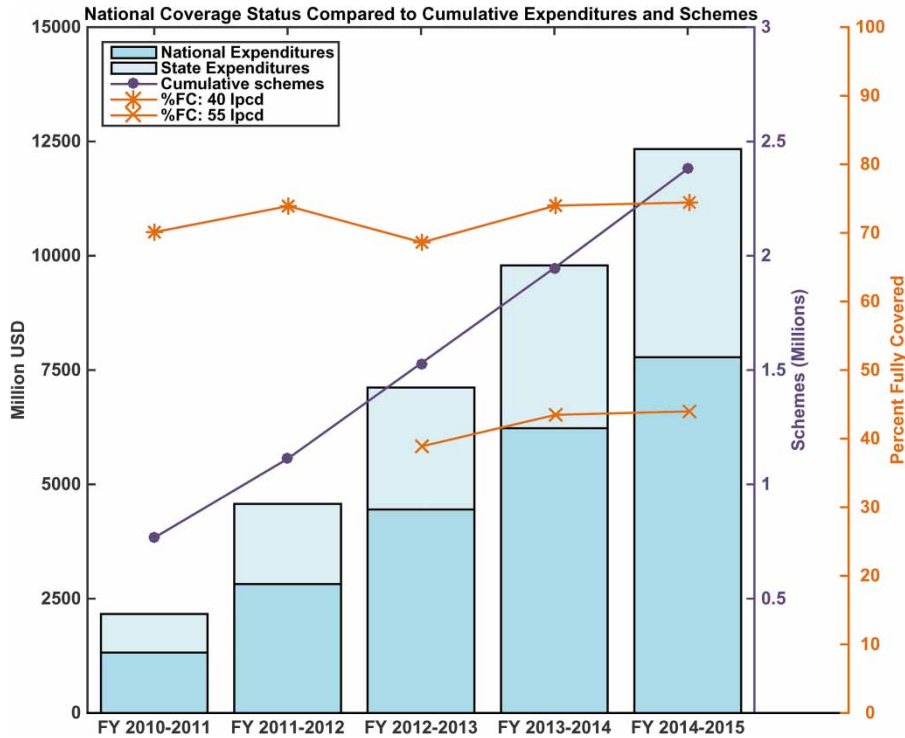


Fig. 1. Percent of fully covered habitations nationwide, cumulative public expenditures, and cumulative schemes from FY 2010–2011 to FY 2014–2015.

by state. The highest levels of coverage ($>90\%$) include states like Gujarat, which have a strong record of rural water expenditure and implementation (Shah *et al.*, 2009). It is curious to observe that some poorer states like Jharkhand and Uttar Pradesh also report very high levels of coverage, which may reflect access to shallow groundwater or may possibly raise data consistency or data quality questions across states. A recent detailed field study of districts in Maharashtra by the Tata Institute of Social Sciences found that IMIS over-represented full coverage by 13% (Sakthivel *et al.*, 2015). Interestingly, the lowest levels of full coverage are reported in north-eastern states and in Kerala, which are different from one another in most respects. Both have relatively high levels of monsoon rainfall, but the former have remote areas of tribal settlement, while the latter has high human development indicators but relatively lower economic growth.

There is greater variance in states' current water coverage at 55 lpcd, and the challenge of meeting the 55 lpcd standard will be greater in all but a few states, such as Gujarat (Figure 3(a) and 3(b)). Interestingly, some economically prosperous states like Maharashtra have only a small proportion of their rural habitations served at this higher standard of coverage, while other poorer states have better coverage. They will need to address the new planning goal now, as well as strategically addressing lagging pockets of water poverty. The very southern, northern, and north-eastern states also have relatively low levels of full coverage at the 55 lpcd level. These very different environmental and cultural contexts raise the question of whether these perimeter states have other common attributes. Tamil Nadu faces pressing water shortages, the north-eastern states remoteness, and Jammu and Kashmir the limited water infrastructure of mountain settlements.

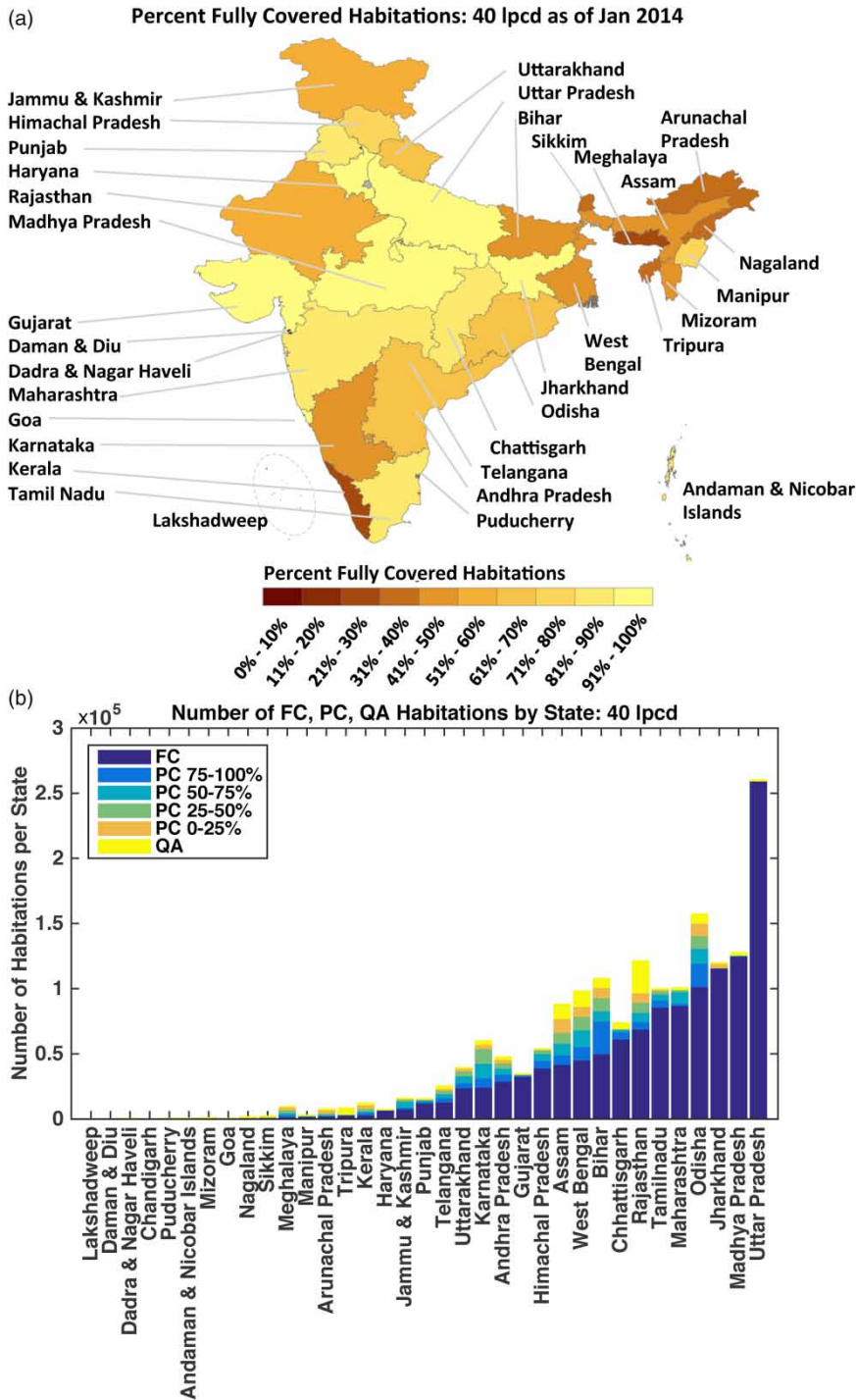


Fig. 2. Water coverage status by state at the 40 lpcd threshold as of January 2014: (a) (top) displays the percentage of habitations that are fully covered; (b) shows the number of habitations in each of the water coverage status categories.

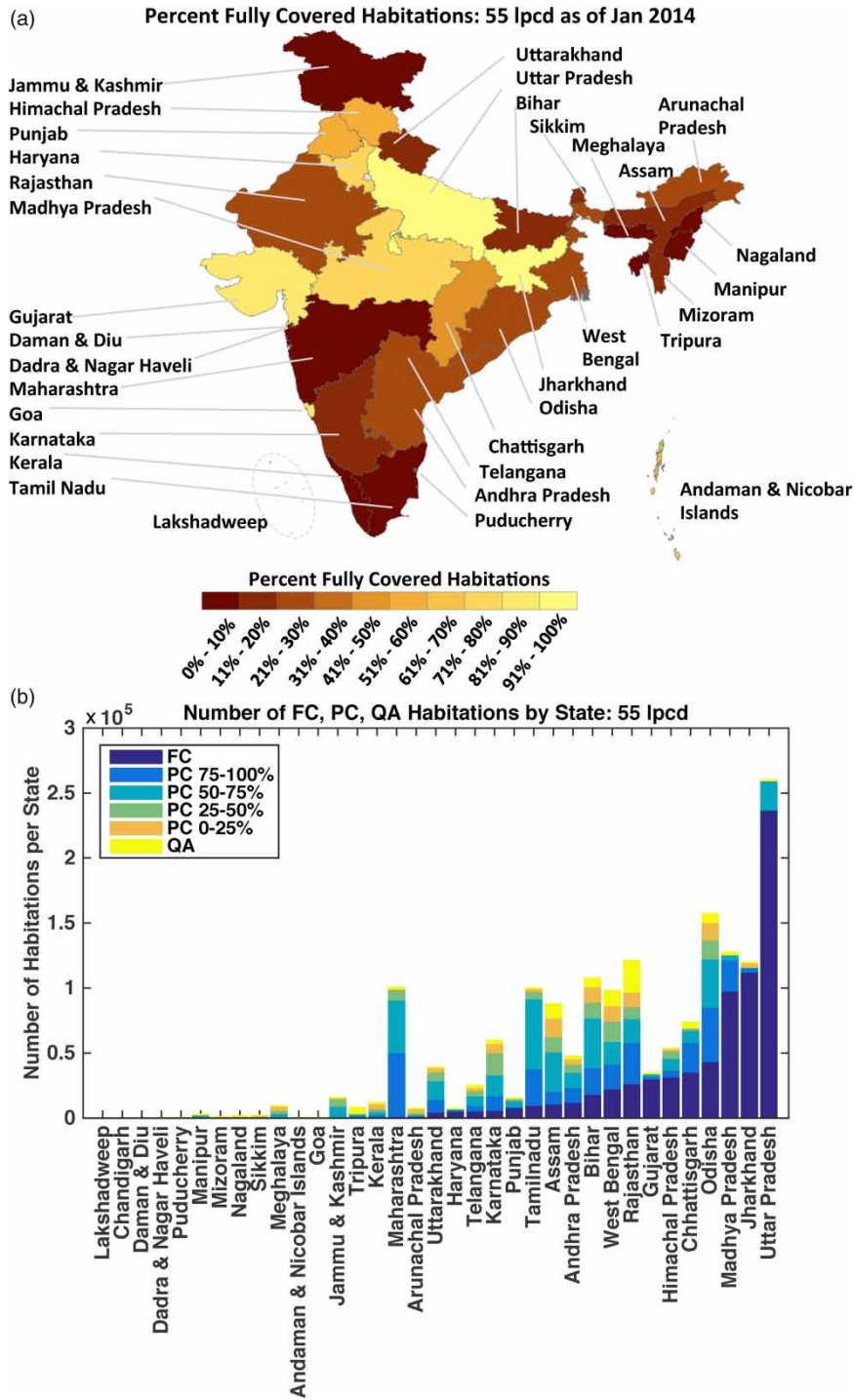


Fig. 3. Water coverage status by state at the 55 lpcd threshold as of January 2014: (a) (top) displays the percentage of habitations that are fully covered; (b) shows the number of habitations in each of the water coverage status categories.

4.2. Water QA habitations

The third major category of drinking water conditions relates to quality, and here we see the greatest challenges for national water policy. Surprisingly, in all but Tripura, water quality standards appear to be met for most habitations (Figure 4) when using quality-affected habitations as the criterion. Figure 5 gives more insight by summarizing the water quality test results of sampled water sources, indicating that the largest raw number of uncontaminated samples was observed in Uttar Pradesh, followed by Madhya Pradesh, and Tamil Nadu. As these states have relatively high levels of poverty and industrialization, these data need to be questioned.

When we look closer at the types of contamination reported by state, four observations may be made. First, the number and percentage of tested sources vary greatly by state. In other words, these data provide a sample rather than a census of water quality. The sampling protocols are not fully specified. Second, and as noted in Figure 4, the proportion of negative test results is very high, perhaps in part because sampling of sources is primarily of groundwater and protected wells. Third, the majority of positive test results involve chemical contamination (e.g. arsenic, fluoride, salinity, and nitrates). Biological contamination reports are surprisingly few in light of sanitation concerns. Finally, the current categorization of habitations as FC, PC, or QA does not allow for failure of both quantity and quality. As the emphasis to date has been on water coverage, water quality has not received the attention needed to achieve health objectives.

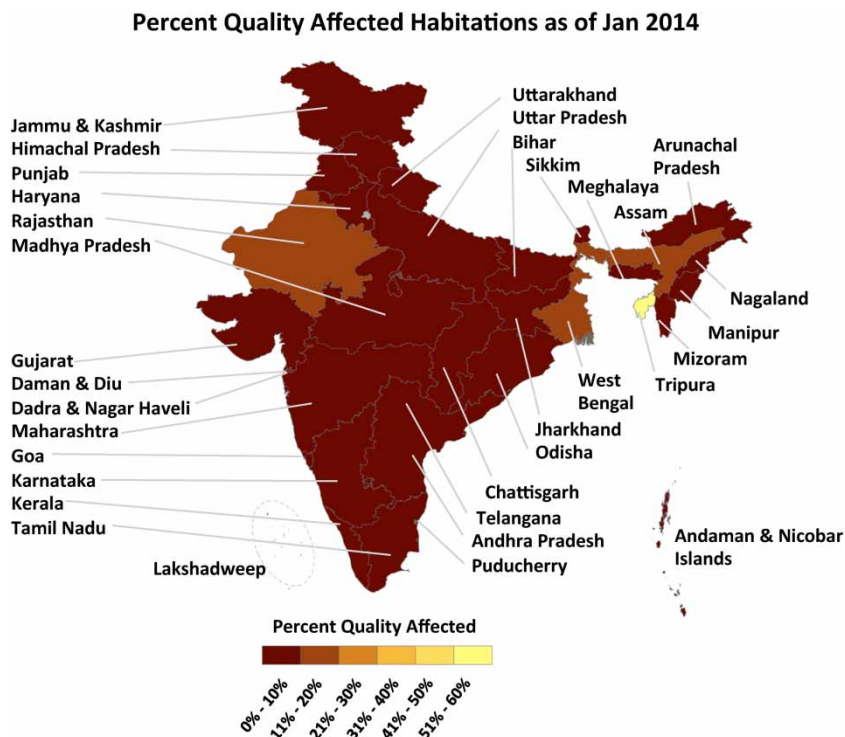


Fig. 4. QA habitations by state (%), 2014.

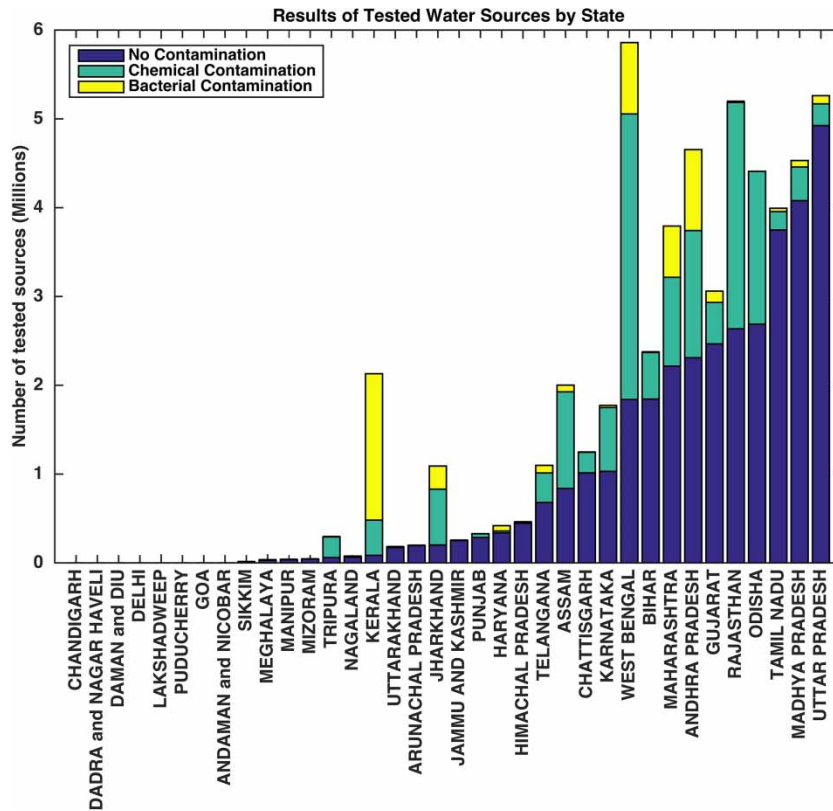


Fig. 5. Water quality test results by state.

Kerala stands out for reporting higher levels of biological contamination per number of wells tested, but overall that variable requires more rigorous examination at the national level. Chemical contamination, e.g. arsenic, fluoride, nitrates, and total dissolved solids (TDS), varies widely across the country, which warrants further analysis of geographic patterns. The Bengal region reports high arsenic contamination rates, as expected. Rajasthan and Karnataka have high TDS, and many agricultural regions of the country have high nitrate-affected habitations. Some of the states with low contamination rates are simply ones with fewer samples tested – such as Himachal Pradesh, Jammu and Kashmir, Arunachal Pradesh, Meghalaya, Mizoram, and Nagaland – so it is possible that states that appear to have low contamination rates are inadequately sampled. These initial observations suggest that water quality testing should be a top priority at the national policy level.

5. The prospects and constraints for explanatory policy analysis and planning

The previous section demonstrated the usefulness of the IMIS database in describing the current state of water access across India. Ideally, IMIS would also allow us to analyze the extent to which government investment in water infrastructure improves water access. In this section, we use regression analysis with the IMIS database to test whether differences in investment, infrastructure, and socio-economic

Table 2. IMIS variables and their expected impact on fully covered (FC) habitations.

Name	Definition	FC impact
Sch. Tot	Number of water infrastructure schemes implemented	Increase
Exp	Total NRDWP government expenditures 2009–2014	Increase
Exp_Ratio	Ratio of exp from central govt to exp from state govt	Increase
DDP	Number of districts in the DDP	Decrease
FC_2009	Number of habitations fully covered in 2009 at 40 lpcd	Increase
LWE	Number of districts affected by left wing extremism	Decrease
Min	Number of blocks with a majority of minority population	Decrease
SC	Population of SC	Decrease
ST	Population of ST	Decrease
Train	Number of members of the VWSC* trained	Increase

*VWSC = Village Water and Sanitation Committee formed by local villagers.

variables can explain differences in state water coverage. We attempt this explanatory analysis, but limitations in the structure of the IMIS database prevent us from developing robust causal inferences. We describe below the methodology, limitations, and recommendations for enhancing database functionality.

The primary goal of regression analysis is to test whether NRDWP programs implemented since 2009 have impacted coverage status. To do this, we examine three key policy components: water infrastructure development, measured by the number of schemes built since 2009; NRDWP expenditures by national and state governments; and community capacity building, measured by the number of persons trained under the NRDWP program. We expect that some of the disparity in coverage status is explained by underlying demographic factors. We therefore also include demographic information that IMIS provides as potential predictors of coverage: Desert Development Program (DDP) blocks, left wing extremism districts, and minority populations of scheduled castes (SC) and scheduled tribes (ST). Furthermore, we expect that some disparity in coverage in 2014 is explained by the coverage status in 2009 at the start of the NRDWP programs. We include coverage status in 2009 in our regression as a control to effectively measure impacts on difference in coverage instead of total coverage. Table 2 defines the predictors considered for independent variables in the regression models, with hypotheses about their impacts on fully covered (FC) status.

5.1. State-level regression analysis

The best method to assess the relationship between the independent variables and water coverage status across India would be to use the full granularity of the IMIS data to develop a habitation-level regression model with data from all 36 states and union territories. However, while data are collected at the habitation level, the web-based public database aggregates those data series up to the district and state levels, making it difficult, if not impossible, to access the raw habitation-level data¹. We

¹ Note that the database aggregates information using unique ID numbers for schemes and sources. This prevents double counting of schemes, investments, etc. when the data are aggregated. This aggregation structure is documented in Novellino (2015) based on conversations with government officials and has been confirmed by the authors through a sample of 550 schemes in five districts.

therefore use state-level data to construct an initial countrywide model. The dependent variable is the proportion of fully covered (FC) habitations (i.e. the number of FC habitations out of the total number of habitations in a state). This state-level approach limits the number of observations in our dataset to 30 (this includes the removal of six states and union territories that have significant data gaps).

We use logistic regression with a binomial formulation, the most common model for dependent variables that are proportions, and develop several models from the variables available in Table 2 using common variable selection methods. However, the analysis of fit for all these models finds that none of them is a good fit; that is, none of them is able to assess the impact of investment and infrastructure variables on full water coverage status. In fact, most of them fail to perform significantly better than a constant model without any predictors. We ruled out typical model formulation problems by using data transformations, removing outliers, and testing alternative model structures instead of logistic regression. This leaves us to conclude that the data are insufficient to parameterize an accurate model. It is likely that some key predictors are missing, e.g. household or per capita income. Data quality issues at the state level of aggregation may also be relevant. Additionally, it's possible there are data reporting problems that did not show up as outliers. It is more likely in India, however, that the 30 state-level observations used here are too small a sample size, with too much variance within states, to fit a strong nationwide model that predicts habitation-level water coverage status.

We therefore conclude first, that the current IMIS database does not, by itself, enable state-level explanations of national water coverage and, second, that the IMIS database should develop increased functionality for national analysis using district-level and, if possible, habitation-level, data to enable explanatory policy analysis at the national level.

5.2. Habitation-level analysis

While the IMIS database does not provide ready access to national habitation-level data on the policy measures we are evaluating, we were able to obtain habitation data from the central headquarters of the IMIS at the National Informatics Center (NIC) on water coverage, the population of SC, the population of ST, and the general (non-SC or -ST) population (Gen Pop) at the habitation level. Given that we expect high SC and ST populations to be predictors of low water coverage status, we now present a preliminary habitation-level regression analysis for a case study district. Building on Novellino's (2015) research in Gujarat, we downloaded habitation-level data on coverage status for Gandhinagar, the capital district of Gujarat. We chose to use the 55 lpcd threshold as there was greater variation across habitations than at the lower threshold.

Table 3. Descriptive statistics for habitations in Gandhinagar district ($n = 496$).

	SC	ST	Gen Pop
Min	0	0	0
Median	83	1	1,176
Mean	128	25	1,945
Max	1,850	251	11,158
Std Dev	167	50	2,122

Gandhinagar district was selected from the state water supply study, in part because it has a diverse population (Table 3). Out of the 496 habitations in the district, 448 of them have an SC population, and 274 have an ST population. Additionally, 76% of the habitations have a fully covered status at the 55 lpcd threshold. The odds of a habitation in Gandhinagar having a fully covered status are about 3.2 to 1.

Correlation analysis shows, as expected, a significant negative correlation between ST and coverage status. Interestingly, there is no significant correlation between SC and coverage status. Additionally, there is a significant positive correlation between Gen Pop and coverage status, indicating that habitations with larger populations are more likely to be fully covered.

5.3. Model fitting

We again use logistic regression, now formulated for binary data. Now each observation is a habitation, and the dependent variable is a binary variable indicating whether or not the habitation is fully covered at the 55 lpcd threshold. Standard transformation analysis led us to use a square root transformation of the independent variables. We test the model fit and confirm that all three independent variables are significant and worthy of inclusion in the model. This process yields the regression model and results in Table 4.

Assessing model fit for a logistic regression model is somewhat more complex than for linear regression. The common interpretation of R^2 for linear regression does not hold in logistic regression (Hilbe, 2009); we use a log-likelihood pseudo- R^2 , which is 0.3575. This indicates a relatively weak model fit that is likely missing some important predictors, as expected. Additionally, standard outlier analysis identified many outliers, and repeating the analysis without the outliers yielded a second model with new outliers. This suggests a problem with the model formulation; most likely additional predictors are needed.

As in the state-level analysis, data availability prevents us from answering key policy questions. The issue here is different from that in the national analysis. We have the granularity in data needed for a district in India, but IMIS provides a small subset of potential socio-economic and institutional predictors. That said, we conclude with fairly high confidence that there is a negative relationship between ST population and FC status, and a positive relationship between population size and FC status in Gandhinagar district. The best regression model developed, which is shown in Table 4, indicates that an

Table 4. Gandhinagar habitation-level regression model and results.

Model:
 $\text{logit}(\text{Status}) \sim 1 + \text{sqrt}(\text{SC Pop}) + \text{sqrt}(\text{ST Pop}) + \text{sqrt}(\text{Gen Pop})$

Estimated Coefficients:

	Estimate	SE	tStat	P-value
Intercept	0.8522	0.31541	2.7018	0.0071
sqrt(SC Pop)	-0.0915	0.0340	-2.6928	0.0073
sqrt(ST Pop)	-0.2736	0.0450	-6.0843	2.35×10^{-9}
sqrt(Gen Pop)	0.0584	0.0105	5.5675	4.26×10^{-9}

496 observations, 492 error degrees of freedom
 Estimated dispersion: 1.25
 F-statistic vs. constant model: 52.4, $P\text{-value} = 2.08 \times 10^{-29}$

increase of 1 in the square root of ST population decreases the odds of being fully covered by 1.31 times, while a unit increase in the square root of the general population increases the odds of being FC by 0.94 times. SC population was not statistically significant in this model, although it did have a significant negative relationship in other models tested. More information is needed to assess the complex local relationships between SC populations and FC status.

6. Conclusions and implications

Five major conclusions for national rural drinking water programs stand out. First, India's investment in an online national rural drinking water database is an important precedent for other countries. Second, it is designed to compile consistent, systematic, transparent, and secure rural water data for policy support. Third, the IMIS database reveals the potential, and current limitations, of a national water database, particularly for data quality control and applied policy analysis. It has yet to be demonstrated how national-, state-, and district-level administrators and water managers actually understand, navigate, and use the large number of tables in the database. Fourth, in descriptive terms, the IMIS database helps monitor advances in national and state water coverage (e.g. toward the 40 and 55 lpcd standards), and related water source, scheme, and sustainability variables. This analysis highlights the need for much greater emphasis on water quality monitoring. Fifth, while IMIS is valuable for descriptive monitoring, our regression analysis experiments showed that it currently has significant limitations for policy analysis. The regression analysis showed that state-level data in the IMIS database, by itself, cannot explain national patterns of full water coverage. Additional socio-economic variables from other databases (e.g. Census of India) could help address this issue. However, we also showed that it is more likely that policy analysis will require district-, block-, and habitation-level observations. The current IMIS database could have greater functionality by providing ready access to district- and block-level data nationwide (*vis-à-vis* for individual states). However, as an annual survey of water supply, the greatest power of the IMIS database will lie in habitation-level regression analysis. This will require greater access to habitation-level data in formats conducive to large-scale regression analysis. In contrast with the national-level models examined here, regression analysis of habitation-level data for the Gandhinagar case study district identified a significant negative relationship between the population of ST and full water coverage, but not between SC and full water coverage. This analysis also indicated that habitation size is positively correlated with water coverage. As might be expected, small systems need strategic emphasis.

The potential for more rigorous and useful policy analysis with the IMIS database thus appears to depend upon: (1) enhanced data access and web interface functionality; (2) ready linkages with other socio-economic databases; and (3) an emphasis on district-, block-, and, above all, habitation-level data.

Acknowledgments

We are grateful to the MIT-Tata Center for Technology and Design and director Dr Robert Stoner for supporting this research. In the Government of Gujarat, Mr Mahesh Singh and colleagues in the Water and Sanitation Management Organization (WASMO) were very helpful. Mr Divyang Waghela of the Tata Foundation water mission offered valuable insights. Ms Seemantinee Sengupta from the

NIC – IMIS provided fundamental support for this research. In Maharashtra, the Department of Water Supply and Sanitation, Groundwater Survey and Development Authority, zilla parishads, and Jalswarajya II project with the World Bank are collaborating on important extensions of this work. Mr J. V. R. Murty offered encouragement and insights on demand management. Architect Surekha Ghogale of the Aga Khan Planning and Building Services, India, and her team supported fieldwork in rural Gujarat. Sean Furey of the RWSN provided useful information on country databases.

References

- Australia New South Wales (2014). NSW Health Drinking Water Database – Drinking Water Quality. Retrieved from: <http://www.health.nsw.gov.au/environment/water/Pages/drinking-water-database.aspx> (accessed 30 June 2015).
- Balasubramaniam, D., Chatterjee, S. & Mustard, D. (2014). Got water? Social divisions and access to public goods in rural India. *Economica* 81, 140–160.
- Biswas, P. K. & Mandal (Mitra), K. (2010). Drinking water in rural India: a study of deficiency, quality and some social implications. *Water Policy* 12(6), 885–897. <http://doi.org/10.2166/wp.2010.005>.
- Cronin, A. A., Prakash, A., Priya, S. & Coates, S. (2014). Water in India: situation and prospects. *Water Policy* 16(3), 425–441. <http://doi.org/10.2166/wp.2014.132>.
- Cronin, A. A. & Thompson, N. (2014). Data and monitoring in the Indian rural water and sanitation sector – a review of current status and proposed ways forward. *Journal of Water, Sanitation and Hygiene for Development* 4(4), 590–603. <http://doi.org/10.2166/washdev.2014.018>.
- Department of Drinking Water Supply (DDWS) (2011). National Rural Drinking Water Program (NRDWP) Strategic Plan 2011–22. Government of India, New Delhi.
- Harrison, M. (1994). *Public Health in British India: Anglo-India Preventive Medicine 1859–1914*. Cambridge University Press, Cambridge.
- Hilbe, J. M. (2009). *Logistic Regression Models*. Chapman & Hall/CRC, Boca Raton, FL, USA.
- James, A. J. (2011). India: lessons for rural water supply: assessing progress towards sustainable service delivery. IRC, The Hague. Retrieved from: <http://www.ircwash.org/sites/default/files/james-2011-india.pdf> (accessed 18 December 2015).
- Khurana, I. & Romit, S. (2009). WaterAid: Drinking water quality in rural India: Issues and approaches. Retrieved from: <http://www.wateraid.org/~media/Publications/drinking-water-quality-rural-india.pdf> (accessed 30 June 2015).
- Lafferty, A. K., Lauer, B., American Water Works Association and Qualseve Benchmarking Program (2005). *Benchmarking Performance Indicators for Water and Wastewater Utilities: Survey Data and Analyses Report*. American Water Works Association, Denver, CO, USA.
- Maupin, M. A., Kenny, J. F., Hutson, S. S., Barber, N. L. & Linsey, K. S. (2014). Estimated use of water in the United States in 2010: U.S. Geological Survey Circular 1405 (Circular) (p. 56). Retrieved from: <http://dx.doi.org/10.3133/cir1405>.
- Novellino, M. (2015). *Analysis of Slipback of Rural Water Supply Systems in India using FIETS Framework and IMIS database – Gujarat Case Study*. Massachusetts Institute of Technology (MIT), Cambridge, MA, USA.
- Prokopy, L. S. (2005). The relationship between participation and project outcomes: evidence from rural water supply projects in India. *World Development* 33(11), 1801–1819. <http://doi.org/10.1016/j.worlddev.2005.07.002>.
- Safe Water Network (2014). Community Safe Water Solutions: India Sector Review. Retrieved from: http://www.safewaternetwork.org/sites/default/files/SWN_India%20Sector%20Review_Sept%202014_Full_Report.pdf (accessed 18 December 2015).
- Sakthivel, R., Sekhar Dhar, N., Godkhe, A. & Gore, G. (2015). Status of Rural Water Supply in Maharashtra. Retrieved from: http://www.indiaenvironmentportal.org.in/files/file/TISS_KRC_Study_Report_Status_of_Water_Supply_in_Maharashtra.pdf (accessed 18 December 2015).
- Satterthwaite, D. (2003). The millennium development goals and urban poverty reduction: great expectations and nonsense statistics. *Environment and Urbanization* 15(2), 179–190.
- Schouten, T. (2015). It all starts with knowing. RWSN Newsletter, 1.
- Shah, T., Gulati, A., Hemant, P., Shreedhar, G. & Jain, R. C. (2009). Secret of Gujarat's Agrarian Miracle after 2000. *Economic and Political Weekly* 44(52), 45–55.

- Shrivastava, B. K. (2013). Mitigation of naturally occurring fluoride in drinking water sources in rural areas in India: an overview. *Journal of Water, Sanitation and Hygiene for Development* 3(3), 467–478. <http://doi.org/10.2166/washdev.2013.107>.
- Srikanth, R. (2009). Challenges of sustainable water quality management in rural India. *Current Science* 97(3), 317–325.
- Ssozi, D. & Danert, K. (2012). *National Monitoring of Rural Water Supplies. How the Government of Uganda did it and Lessons for other Countries*. RWSN, St. Gallen, Switzerland.
- UNECE (2014). Monitoring of a Water Goal: An Expanded Global Monitoring Initiative for Wastewater, Water Quality and Water Resource Management. Retrieved from: http://www.unece.org/fileadmin/DAM/env/water/meetings/workshop_target_setting/Regional_Workshop/3.3.K.Medlicott_Post2015Monitoring_Wastewater.pdf (accessed 18 December 2015).
- UNICEF (2013). Water in India: Situation and Prospects. Retrieved from: <http://www.im4change.org/docs/105water-in-india-report.pdf> (accessed 18 December 2015).
- US Census Bureau (2009). Water and Sewage Systems – The 2012 Statistical Abstract – U.S. Census Bureau. Retrieved from: http://www.census.gov/compendia/statab/cats/energy_utilities/water_and_sewage_systems.html (accessed 30 June 2015).
- US Environmental Protection Agency (EPA) (2015). Safe Drinking Water Information System (SDWIS) Search [Data & Tools]. Retrieved from: <http://www.epa.gov/enviro/facts/sdwis/search.html> (accessed 30 June 2015).
- Water Point Mapping (2015). Retrieved from: <http://www.waterpointmapping.org/> (accessed 30 June 2015).
- WaterAid (2011). *India Country Strategy 2011–2016*. WaterAid, New Delhi.
- Wescoat Jr, J. L. (2014). Searching for comparative international water research: urban and rural water conservation research in India and the United States. *Water Alternatives* 7, 199–219.
- Wescoat Jr, J. L., Theobald, R. & Headington, L. (2013). Water and poverty in the United States. *Encyclopedia of Earth*. Retrieved from: <http://www.eoearth.org/view/article/156916/> (accessed 18 December 2015).
- WHO and UNICEF Joint Monitoring Programme (2015). Improved and unimproved water and sanitation facilities. Retrieved from: <http://www.wssinfo.org/definitions-methods/watsan-categories/> (accessed 30 June 2015).
- WHO and UNICEF (2006). Core questions on drinking water and sanitation for household surveys. Retrieved from: <http://apps.who.int/iris/handle/10665/43489> (accessed 18 December 2015).
- World Bank (2014). India – Maharashtra Rural Water Supply and Sanitation Project (RWS III). Retrieved from: <http://www.worldbank.org/en/news/loans-credits/2014/03/12/india-maharashtra-rural-water-supply-sanitation> (accessed 30 June 2015).
- World Bank (2015). IBNET. Retrieved from: <https://www.ib-net.org/> (accessed 18 December 2015).

Received 18 July 2015; accepted in revised form 25 November 2015. Available online 1 February 2016