Informing the Water and Sanitation Sector Policy: Case Study of an Impact Evaluation Study of Water Supply, Sanitation and Hygiene Interventions in Rural Maharashtra, India

Subhrendu K. Pattanayak
Christine Poulos
Kelly M. Wendland (Conservation International)
Sumeet R. Patil
Jui-Chen Yang
Richard K. Kwok
Catherine G. Corey (Center for Studying Health System Change)

Working Paper 06_04 February 2007

RTI International's* Environment and Natural Resource Economics Program circulates these working papers to convey information and entertain discussion. The papers may not have undergone formal peer review. Contact Subhrendu K. Pattanayak (subhrendu@rti.org) if you have any questions or concerns. <www.rti.org/enrepaper/>

Informing the Water and Sanitation Sector Policy: Case Study of an Impact Evaluation Study of Water Supply, Sanitation and Hygiene Interventions in Rural Maharashtra, India

Subhrendu K. Pattanayak, Christine Poulos, Kelly M. Wendland,
Sumeet R. Patil, Jui-Chen Yang, Richard K. Kwok, and Catharine G. Corey

1. Introduction

Sustainable and equitable access to safe water and adequate sanitation are widely acknowledged as important development goals (e.g., UN, 2005; UNDP, 2006). Yet, to date we have few or no rigorous scientific impact evaluations showing that water supply and sanitation (WSS) policies are effective in delivering many of the desired outcomes (Fewtrell and Colford, 2005). Impact evaluation measures the impacts of the program on individuals, households, and communities, and determines whether the program *caused these impacts* (Baker 2000; WB-OED 2004). An impact evaluation relies on control or comparison groups, as well as a number of statistical and econometric techniques to determine what would have happened in the absence of the program – this is known as the *counterfactual*.

We have speculated on several potential reasons for the paucity of rigorous impact evaluations in the WSS sector elsewhere (Poulos et al. 2006), but we focus here on three reason, all of which are features of WSS policies and interventions.

First, mechanisms to achieve these goals are broad and varied in terms of the types of services (water supply, water quality, sanitation, sewerage, and hygiene); the setting (urban, peri-urban, rural); and the typology of delivery (public intervention, private interventions, decentralized delivery, expansion or rehabilitation). While these complex interventions call for carefully designed evaluation studies, most previous impact evaluations have had insufficient designs for measuring program impacts and/or for measuring the full range of impacts. One exception is the UNDP (2006) report linking adequate and safe water and sanitation to Millennium Development Goals (MDG) related to child mortality, maternal health, poverty and hunger, education, gender equity, women empowerment, combating HIV, malaria and other diseases, and environmental sustainability.

Second, decentralized and community-level projects – particularly those that are community-demand-driven (CDD)[1] or community participation based – are an important and growing class of development projects. The combination of voluntary participation in self-selected interventions by communities and targeted provision by program administrators increases the difficulty of identifying an appropriate control group.

[1] Mansuri and Rao (2004) define community-demand-driven development projects as those in which communities have direct control over key project decisions, including management of investment funds.

Third, the breadth of effects of WSS policies, which range from greater efficiency in the sector at the national level to improved health at the individual level, raises two challenges for impact evaluation. The first is that the engineering and fiscal outputs that are tracked in a Management Information System (MIS) by many projects yield little information on the effects of the program on poverty reduction – the underlying goal of development processes. The second challenge is that assessing these broad impacts requires a thorough and thoughtful approach to study design. Most impact evaluations of WSS programs focus primarily on health or a limited set of outcomes, and therefore, do not collect enough data to evaluate other impacts such as increased educational opportunities, improved rural livelihoods, or gender equity.

Most rigorous impact evaluations in the sector are quasi-experimental designs due to the fact that the features of water and sanitation programs often make randomized designs infeasible.[2] In particular, many WSS programs are both targeted by program administrators and/or driven by community demand, making random assignment for a study inappropriate.

Our purpose in this paper is to show how quasi-experimental impact evaluation tools can be used to address these three key evaluation challenges in the WSS sector. These tools are illustrated by describing a case study of the design and implementation of a quasi-experimental evaluation of a public WSS intervention in Maharashtra, India.

First, through careful sample design, our quasi-experimental design is capable of measuring the relative contributions of multiple project components (e.g., water supply, sanitation, information) to multiple project effects (including increased access to services, health effects, incomes, expenditures, and gender equity, among others).. Whereas randomizing such a complex mix of project components would be difficult to implement and monitor (and impossible when communities select project components), quasi-experimental designs can be designed around multi-sectoral and multi-dimensional programs without constraining or influencing the implementation.

Second, the quasi-experimental design addresses the selection bias arising from targeted or CDD WSS program by employing propensity score matching (PSM). We use 'pre-intervention' PSM matching to identify a control group that is comparable across multiple variables with the treatment or the program group. This is a significant deviation from most evaluation studies that rely on post-intervention PSM using only two to three matching variables. The weakness of post-intervention PSM is its reliance on secondary data for pre-intervention (or baseline) measures, or even the complete lack of baseline data altogether when the relevant data are not available from secondary data. Our evaluation uses pre-intervention matching on over 20 variables, which permits the identification of controls prior to the intervention and allows the us to measure the full range of impacts, as well as controls, intervening and mediating variables, and potential confounders. This has only been done in one other impact evaluation study that we are aware of (Lechner et al., 2000).

---

[2] Information and education treatments are exceptions and there have been two recent randomized impact evaluations of information and education treatments in India (Dickinson et al. 2007; Jalan and Somanathan 2004).

Third, our study collects pre-intervention and post-intervention data on communities, households, and individuals, as well as program implementation using a variety of qualitative and quantitative methods to triangulate our measures of impacts and controls. This permits a thorough understanding of the program as well as measurement of a broad range of impacts. Previous impact evaluations in WSS have focused primarily on health.[3]

This paper is organized as follows. The next section takes a stock policy and implementation issues and impact evaluations in the WSS sector. The third section describes the quasi experimental design of our impact evaluation case study of a CDD water, sanitation, and hygiene (WSH) program being implemented in rural parts Maharashtra state in India. The fourth section presents our conclusions about the role of impact evaluation in addressing key issues in WSS program implementation and policy.

2. Background

This section reviews the impact evaluation literature from the WSS sector in two ways. First, we describe the WSS policy landscape and review studies that provide evidence on the effectiveness by policy or type of delivery. Second, we review the evidence by type of water and sanitation impact.

2.1 Evaluations by WSS Policies and Programs

Increasingly, donors (including the World Bank) and agencies have broadened their objectives from a narrow focus on physical infrastructure to the provision of financially sustainable service provision. There are three types programs or policies, referred to as 'sector reforms', are predominantly used to achieve the broader objectives of increased financial viability and improved the institutional performance of the WSS sector: (1) improving operator performance; (2) service provision by the private sector providers (PSP) or small-scale independent providers; and (3) decentralized delivery, typically relying on community demand, participation and management. Increasingly, another type of delivery, public-private partnerships (PPPs), is being used to deliver WSS services, as well as interventions in other sectors.

The first type of sector reform focuses on helping utilities reduce costs and increase revenues to become financially viable, thereby improving and extending service delivery. Despite interesting case studies (Drees et al., 2004; Mumssen, 2004), there are no known impact evaluations of these types these reforms.

The second type of policy, PSP (large-scale projects in which one or more corporate entities, with private equity, assume operating risk and/or development under a license or contract), is being used in a growing number of developing countries. We are aware of two rigorous evaluations of PSP, though there remains concern that the poor are not served by these providers. Galiani et al. (2005) evaluate the impacts of Argentina's privatization of water services on access and health using historical mortality data for municipalities with and without privatized water

[3] For examples of rigorous evaluations of health impacts of WSS, see Fewtrell et al. 2005, Kaufmann 2005, Esrey et al. 1991, Esrey 1996, Curtis and Cairncross 2003.

services. Using a number of difference-in-difference (DID) approaches and using PSM, the authors find that privatization decreased child mortality rates by 5 to 7 percent. Clarke et al. (2004) match treatment municipalities to control municipalities on the basis of population size and use household level data on access to piped water from Bolivia, Brazil, and Argentina. They conclude that privatization does not increase access since access increases equally in both treatment and control cities.

With one-third of the population of Africa and Asia living in towns with populations that range between 2,000 and 200,000 people, the third type of policy – decentralized water supply and sanitation – is fundamental to economic growth and achievement of the Millennium Development Goals. In rural areas, CDD decentralized delivery often translates into putting the community front and center of the planning, design, implementation, and operations process and replacing career bureaucrats with qualified professionals and technocrats in guiding this process. The CDD philosophy is typically captured in reflecting the needs of community through a participatory approach, decentralized delivery, cost sharing (typically 10-30% of capital, and 100% of O&M), and a strong component of local institutional strengthening. This policy has been the subject of several impact evaluations.

Rawlings et al. (2004) summarize the quasi-experimental impact evaluations of social investment funds (SIF) in WSS in Armenia, Bolivia, Honduras, Nicaragua, and Peru. While these studies find improved access to water (measured by decreases in distance to water and time spent collecting water), the health impacts of water supply improvements (measured by child and infant mortality, lost working time, and less stunting) are measured at only a couple of sites. Further, Mansuri and Rao (2004) review studies that evaluate community-level economic development programs in a number of sectors, including WSH, labor, agriculture, and others, but cannot establish a causal relationship between any outcome and the participatory elements of the projects.

The fourth type of the delivery methods that are becoming increasingly prevalent in India (McKenzie and Ray 2005) and elsewhere are partnerships between the public and private sector. PPPs are employed by international organizations to ensure the provision and quality of public health services in developing countries (Buse and Waxman 2001; WSP 2003). To our knowledge, there are almost no rigorous studies of the effectiveness of PPPs in achieving public health objectives (WSP 2003), and certainly no evaluations of PPPs in the WSS sector.

Finally, WSS interventions may be delivered at the community or household level. For instance, public utilities provide services at the community level by increasing the number of sources or the quality of water for all members of the community. Alternatively, some services, particularly water treatment and hygiene education, may be provided at the household level. Indeed, many PPPs provide household level services and CDD provide mainly community level services. There have been several studies that have compared the effectiveness of community-level WSH interventions (e.g., protection or treatment of water sources) to household-level interventions (e.g., point-of-use water treatment), but there remain several unanswered questions. In particular, while the most recent systematic reviews of the epidemiologic literature suggest water treatment at the household level is more effective in preventing enteric disease than improvements at the source, the effectiveness varies by the setting and some studies have methodological flaws that limit their comparability. Fewtrell et al. (2005) and Clasen et al. (2006) find that household water treatment is more effective in reducing diarrhea than community water source treatment. However, Clasen et al. note that these finds are limited to rural settings – household water treatment systems had no effect on health outcomes in peri-urban, urban, and refugee settings. On the other hand, Van Derslice and Briscoe (1993) systematically compare water contamination at home and source and find that improving source water is more effective in decreasing diarrhea rate.

Also, there are important differences in the methodological quality of the studies, their design, duration of follow-up, participant compliance, and prevalence of specific pathogens. First, household studies tended to be randomized, while all source-based interventions were quasi-randomized, introducing potential bias in the comparison of these studies. Second, the duration of follow-up was substantially longer for the source-based interventions, relative to the household studies. With seasonality a recognized determinant of diarrhea incidence, failure to carry-out a study with a sufficient follow-up period may misrepresent the disease burden in the population, influencing the measure of the impact. Third, the studies do not account for household hygiene and water safety behaviors that may affect water quality (Patil and Pattanayak, 2007). Further, courtesy bias or Hawthorne bias (e.g., behavior is affected by the study) may have arisen in the household interventions, since most were not blinded, leading to the over or understatement of the effectiveness of interventions. In concluding, Clasen et al. (2006) recommend that given these differences, comparing various approaches to improving water quality with extended follow-up should be undertaken to more comprehensively assess the merits of household versus community interventions to reduce the burden of diarrheal diseases. Therefore, there remain several important questions about the effectiveness of the main types of policies for delivering WSS services, including sector reforms, PPPs, and community vs. household interventions.

2.2 Evaluation by WSS sector Outputs, Outcomes, and Impacts

The previous section reviews the evidence on the effectiveness of various WSS policies. But, WSS policy is not an end in itself. It is a means to provide a basis for expanding access and improving the quality of service. WSS policies provide *resources* (i.e., available financial, human, social, and institutional capital) that support downstream *activities* (i.e., actions and processes carried out by the program) and *outputs* (i.e., any direct product of program activities that program providers have direct control over) in the sector, including the construction, expansion, and/or rehabilitation of water supply, sanitation, and sewerage infrastructure. Outputs may also include investments in education aimed at changing personal hygiene behaviors and/or institutional change to affect the behavior (operation and management) of organizations, including utilities, communities, and cooperatives.

We now describe evaluations that focus on these downstream effects of WSS policies. We classify these downstream effects as *outputs*, *outcomes*, or *impacts*. Figure 1 illustrates and defines these effects.

Figure 1. Model of a Generic WSS Program

	Program Resources	è	**Program Activities**	è	**Program Outputs**	è	**Program Outcomes**	è	**Program Impacts**
Definition	financial, human, social, and institutional capital	è	actions and processes	è	type of products and levels of service delivered to participants	è	changes in behaviors, knowledge, and actions among participants as a result of the program	è	fundamental changes are improvements in beneficiaries' well-being
Example	Program budget Program management team	è	Funds disbursed for infrastructure improvements, technical guidance provided	è	Increased water supply capacity Number of new latrines	è	Improved access and use of water supply and sanitation infrastructure	è	Improvements in Health Education Gender equity Income

WSS *outputs* are simply the types of products and levels of service under the direct control of program providers, whether they are public sector, private sector, or community organizations. (Bosch et al. 2000; Prenusshi et al. 2000) We can classify WSS outputs into four categories: (1) water supply (quantity); (2) water quality; (3) individual sanitation; and (4) environmental sanitation. Outputs also include hygiene information and education campaigns for beneficiaries, training for sector staff, and improvements in utility and sector management (e.g., appropriate bookkeeping practices, improved billing, improved accountability, number of concession contracts awarded).

Program *outcomes* are the changes in behaviors, knowledge, and actions among participants as a result of the program. The primary outcome of WSS interventions is access or use. The indicators of access depend on the outputs of the program. Some indicators, such as the average distance from beneficiaries' homes to a water source, measure the availability of services. Other indicators, such as the liters consumed per capita per day, the number of hours of service, and the quality of drinking water, measure the quality of services. Typically, outcomes are realized at the household or individual levels.

WSS *impacts* are the fundamental change experienced by program beneficiaries as a result of the program. Bosch et al. (2000) categorize impacts into four groups: (1) health improvements; (2) education; (3) gender and social inclusion; and (4) income/consumption increases. To our knowledge, no study has demonstrated that any WSS program or policy has generated all four impacts.

In the simplest terms, rigorous impact evaluation is primarily concerned with measuring outcomes and impacts. However, it is important to understand what happens on the left side, i.e. the resources, activities and outputs. For example, it is possible that a program's intended impacts are not met because the program activities were not implemented as planned. These issues are analyzed in a more qualitative process evaluation which supports impact evaluation and together they define "program evaluation". Also, evaluating WSH interventions obviously necessitates adequacy of these interventions as intended. Adequacy surveys help assess the level of adequacy of the outputs and outcomes to not only schedule an appropriate followup survey but also to monitor, evaluate (and if desired, inform) project implementation. Patil et al. (2006) describe method and present a case study of an adequacy survey. Process evaluation and adequacy surveys are not a focus of this paper.

While Bosch et al. (2000) identify four categories of WSH impacts, the health impacts are studied most frequently. A small number of studies have measured changes in the household's costs of collecting, storing and treating water, as well as the income losses due to water-borne and water-washed illnesses (Pattanayak et al. 2005b). There have been no rigorous studies of the education, gender and social inclusion, and poverty reduction impacts of WSH interventions.

The studies of health impacts include a recent meta-analysis commissioned by the World Bank on rural WSH impacts (Fewtrell et al., 2005), several other reviews (Curtis and Cairncross, 2003; Esrey et al., 1990, 1991; and Huttly et al., 1997), and papers by Young and Briscoe (1987), Baltazar and Solon (1989), Clemens and Stanton (1987), Cousens et al. (1990), Daniels et al. (1990), Galiani et al. (2005), and Han and Moe (1990). There are several studies focused on child health in India, including Jalan and Ravallion (2003), Van der Klaauw and Wang (2003) and Dasgupta (2004).

In general, the epidemiologic studies report effect measures in terms of relative risks or odds ratios while the economic studies tend to evaluate marginal impacts associated with the WSH interventions. WSH interventions that have been studied can be classified into two types: infrastructure services (e.g., increase in access to public handpumps/taps, water quantity and quality, and provision of public latrines), and behavioral changes (e.g., hand washing practices, safe disposal of feces, safe water and food storage, and water treatment). The impact of these interventions has tended to range from approximately 15 to 50 percent reduction in diarrhea morbidity. Covariates considered in these impact evaluations include child characteristics (e.g., gender, age, weight, race, and immunization), maternal characteristics (e.g., age, education, marital status, pregnancy history, ante- and postnatal medical care, and breast-feeding practices), household socio-economics (e.g., household income, income assistance, asset holdings, employment, location, health service utilization, family size, number of children under 5, electricity, cooking fuel, number of rooms, and presence of domestic animals/pets), and community characteristics (e.g., population density, rainfall, and location). Fewtrell et al.'s (2005) meta-analysis and our own literature review identified two broad areas of weakness in the evaluation studies (Pattanayak et al. 2005a).

First, in spite of publications emphasizing the need to improve research quality (Blum and Feachem, 1983; Esrey and Habicht, 1986; Fewtrell and Colford, 2004), 50 percent of the evaluation studies Fewtrell et al. reviewed don't meet quality standards. In addition, Fewtrell and Colford suspect publication bias. Our own literature review also confirmed the lack of scientific rigor in evaluations, particularly with respect to (a) accounting for baseline diarrhea rates and pre-intervention behaviors, (b) the inclusion of control groups, (c) explicit examination and control for confounders, and (d) detailed reporting and presentation of results. Fewtrell and Colford (2004) identify only two high quality evaluations of the health impacts of sanitation projects.

The second reason, related to the first, is that there have been almost no studies of the impacts of decentralized and CDD programs. The design of these programs results in heterogeneity across communities. This introduces challenges to the research design, discussed in Section 2.4, that have not been met in previous studies.

2.3 Evaluations of WSH Interventions in India

McKenzie and Ray (2005) examine water supply and quality conditions in India and survey the range of water supply technologies and institutional arrangements, and conclude, as we have, that there are few peer-reviewed studies evaluating drinking water interventions in rural India. Pattanayak et al.'s ongoing evaluations of public WSH programs in Maharashtra and Orissa are exceptions (described in Corey et al., 2007; Dickinson et al., 2006; Dickinson et al., 2007; Patil et al., 2006; Pattanayak et al., 2006; Patil and Pattanayak, 2007). McKenzie and Ray also find that previous case studies "tend to be brief, optimistic, and not easily comparable to one another."

We also review several recent studies relying on the data from large scale national surveys in India. These studies don't find health impacts of public water supplies, which is the main mechanism of water delivery in rural India (Bonilla-Chacin and Hammer, 1999; Hughes et al., 2001; Wang, 2002; Jalan and Ravallion, 2003). To some extent these results can be explained by the inability to account for water contamination during handling and home storage (Jensen et al., 2002; Sobsey, 2002), or for insufficient flowing water and therefore hand washing (Curtis and Cairncross, 2003). It is also possible that the general purpose national surveys cannot reveal real health benefits of public water supply because their categorization of water supply types is too broad (ignoring issues related to water quality at the point of use, maintenance and proximity of stand posts) and their design does not allow for adequate modeling of all relevant co-factors (e.g., culture, diet, geography, socio-economic status, governance) that affect the linkages between WSS and health. In general, the cross-sectional, non-experimental and non-specific nature of national survey data precludes their use in estimating the causal impact of WSH services. Nor can these data be used to empirically illustrate the causal chain from the investment in WSH programs, through activities, outputs, outcomes, and ultimately to impacts.

Dickinson et al. (2007) describe the results of the one of the first rigorous impact evaluations completed in India. The study used a randomized design to evaluate the impacts of a community-level information, education, and communication (IEC) campaign (the intervention or the program and process) in rural Orissa state on the household-level decisions to build (the output) and use individual household latrines (IHLs) (the outcome), as well as on the diarrhea morbidity (the health impact). Using a DID estimator, they find that the IEC campaign substantially increased IHL uptake by about 30%. This increase in IHL, in turn, may have had a slight impact on child diarrhea impacts, although these results are less robust to different model specifications. The follow-up survey was implemented soon after the intervention was complete, so it is expected that diarrhea rates will continue to decline over time. This hypothesis will be tested using data collected in future survey rounds.

2.4 Why are there so few impact evaluations in WSS?

This literature review has identified various gaps in evaluations in the WSS sector. In their document on guidelines for conducting impact evaluations in the WSS sector, Poulos et al. (2006) speculate on several reasons that there are few rigorous evaluations in the sector.

Several reasons are not unique to the WSS sector and are unrelated to features of WSS policies, and Ferraro and Pattanayak (2006) discuss these in detail. For instance, one usually needs a remarkable combination of political will, a strong commitment to transparency, and a strong ethic of accountability to conduct a well-designed evaluation. Second, WSS program staff may be unaware of state-of-the-art empirical program evaluation techniques and the biases in current analyses. Third, many correctly believe that rigorous evaluations are expensive and but err in considering these evaluations as “non-essential” investments. Fourth, researchers and practitioners in other policy fields have demonstrated that randomized experimental methods can be implemented in the context of small pilot programs or policies that are phased in over time. However, the randomized methods are often fundamentally non-applicable to CDD projects where self-selection and self-management are essential features of the interventions. Fifth, many WSS project implementation cycles are short but even the “immediate” impacts of the project will largely be realized after the project ends. The evaluation “lessons” from these immediate and other long term impacts will accrue to the global community not to the evaluated project per se.

Some of the reasons for a lack of impact evaluations in the sector may be traced back to features of WSS policies and interventions. Similar issues may arise in other sectors as well, but the key point is that the structure of policies makes impact evaluation challenging. We focus on three of these reasons here.

First, the mechanisms used to achieve WSS goals are broad and varied in terms of the types of services (water supply, water quality, sanitation, sewerage, and hygiene); the setting (urban, peri-urban, rural); and the typology of delivery (public intervention, private interventions, decentralized delivery, expansion or rehabilitation). In addition, there is a wide spectrum of possible socio-cultural, economic, environmental, political and legal conditions in which services are delivered. The complexity of water and sanitation increases the challenges facing the impact evaluation researcher. Even the objectives of the WSS programs may be obscure as a diversity of donors and practitioners from different sectors may not agree on a set of explicit objectives. Given this complexity, some previous impact evaluations either have insufficient controls, no baseline measures, and/or insufficient sample size and variability to identify program impacts. Further, the type of delivery (e.g., public, private, or public-private partnership) is likely to bear on the effectiveness of programs, yet the type of delivery is not varied within programs. Thus, evidence on the effect of type of delivery on program outcomes is practically absent.

An issue related to the heterogeneity in services, delivery, and contexts is that the lessons of rigorous impact evaluations of water and sanitation programs may have limited generalizability. In Africa, for example, there is a greater emphasis on installing and rehabilitating infrastructure in order to provide access to safe water from a WSS system. The current level of services tends to be lower than they are in, say, Latin America, where many people have access to the water supply system. In these areas, the quality of service (in terms of regularity, pressure, and water quality) is the focus of WSS improvements. Consequently, there is a need for more local evaluation as well as evaluations that control for important contextual features and have increased generalizability. However, as MacKenzie and Ray (2005) point out, even the studies in the same country may not be generalizable because the issues are completely different for rural and urban settings.

Second, the increasing use of decentralized community-level projects relies on a combination of voluntary participation in self-selected interventions by communities and targeted provision by program administrators. Under these conditions, there are likely to be systematic differences between treatment and control units and this increases the threat of 'selection bias'. Selection bias is also a threat when projects are centralized, since interventions often sequenced and targeted to villages with considerations to equity and political needs.
Third, because of the breadth of WSS impacts, from greater efficiency in utilities to, for example, social and gender inclusion, researchers must carefully select indicators that are highly correlated with the program objectives. For example, engineering and fiscal measures in MIS yields little information on the effects the program on poverty reduction. Also, WSS impact evaluation requires detailed multi-level data. In other fields of policy analysis, researchers have longstanding national surveys and historical relationships with government agencies and field practitioners that generate substantial datasets for research. Quality and quantity of water and sanitation and hygiene information are typically excluded from these data sets. Thus, substantial primary data collection is likely to be required for impacts evaluations in WSS at least in the beginning phases of the global research.

2.5 Quasi-experimental Evaluation

The following three features of quasi-experimental research and impact evaluation can address above challenges in evaluations in WSS sector.
First, even with complex–multi-dimensional–multi-sectoral policies, we can use a thorough understanding of program implementation to design the evaluation study in terms of the sample and the data collection strategy in order to vary (or isolate) the various dimensions of policy interventions in simple ways across space and time.
Second, we can address the threat of selection bias with the use PSM and DID. PSM addresses selection bias due to observable characteristics by matching intervention communities to observationally similar non-intervention communities. DID estimates accounts for selection bias due to unobservable time-invariant differences between intervention and non-intervention communities.
Third, we can collect quantitative and qualitative data at multiple levels to measure program impacts, as well as activities, outputs, and outcomes. We can use quantitative structured surveys to collect community, household, institutional. We can support the analysis using this data with qualitative insights from process evaluation and adequacy surveys. As Patil et al. (2006) demonstrate in the case of adequacy surveys, triangulating quantitative data with qualitative data on program implementation is critical to explain program implementation process and ascertain the adequacy of WSH interventions.

3. Case Study

This section describes a quasi-experimental impact evaluation design that addresses the key challenges in evaluating WSS policies and interventions. The focus of the evaluation is the WSH interventions provided under aegis of a public WSS program in India. This program is an example of the approaches to delivering water-, sanitation-, and health-related services the Government of India (GoI) is using to reach its “national planning objectives” related to universal access to water and reduction in child mortality rate that exceed those under the Millenium Development Goals. For example, the GoI has set up an ambitious target to halve the number of people without access to safe water and basic sanitation by 2015 (GoI, 2002). In the Indian state of Maharashtra, the World Bank (TWB) supported Jalswarajya Project is an example of a cross-sectoral community-driven approach that will provide WSH related infrastructure services. The study is designed to evaluating the health impacts of Jalswarajya in terms of diarrhea morbidity in children. Before presenting the design of the evaluation study, we describe the background on conditions in Maharashtra and the Jalswarajya program.

3.1 Study Area

In an effort to redress the current water supply and sanitation situation, the Government of Maharashtra (GoM) launched the Jalswarajya Project with support from TWB. The state of Maharashtra in western India is among the most developed and prosperous states in India, with literacy rates of 80%, annual per capita income of US$ 539[4] and a variety of economic activities. About half the population is involved in agriculture (which is low by Indian standards), with 29% of the population being cultivators and 26% being agricultural laborers. Maharashtra is also one of the largest states in India with a population of 97 million or 19.6 million households. The average rainfall is 1,180 mm (although this conceals great variations across the state which has drought-prone areas in the center and east), with 85% of rains occurring during the monsoons (June-October), and a temperature range of 54-106°F. Approximately 58% of the population lives in rural areas and about 20% of the population belongs to scheduled castes and tribes (SC/ST).

The infant mortality rate (IMR) for Maharashtra is 44 deaths per 1000 live births and the under 5 mortality rate (U5MR) is 58 per 1000. However, in rural Maharashtra mortality rates are higher. IMR is 51 deaths per 1000 live births and U5MR is 68 per 1000. Twenty-three percent of the children under 3 suffer from diarrhea (1998-1999 National Family Health Survey [NFHS]). The most recent data on rural water supply (1998-1999 NFHS) shows the following service situation: 23% of the rural population has a household connection to the piped network, 25% rely on public standpipe, 23% use a borehole, and 2% rely on protected dug well. Meanwhile over 85% have no sanitation and there is little or no treatment of water at the home. As these statistics suggest, minimal and poor quality water supply and non-existent sanitation services could be contributing to water-washed diseases such as diarrhea.

[4] All monetary values are expressed in terms of 2004 US Dollars.

3.2 Intervention: Jalswarajya

The overall objectives of Jalswarajya are to increase access to rural drinking water and sanitation services and to institutionalize decentralized delivery of rural water supply and sanitation services (RWSS) by local governments (RSPMU, 2004). Jalswarajya relies on voluntary participation by communities, wherein communities select water supply and sanitation services they want from a menu of options, and targeted provision of RWSS by project administrators.

The program is being be implemented from 2003 to 2009 in 2,800 villages from 26 districts. To address the shortcomings of previous programs, the Jalswarajya approach is based on four principles: (1) the project is community demand driven; (2) community share the cost of the projects – 10% of the capital costs and 100% of O&M costs; (3) decision-making is decentralized and made by the community (not by any government or political institution); and (4) sustainable capacity development of local institutions, including traditional institutions (e.g., Gram Panchayats) and new institutions (Village Water and Sanitation Committees [VWSC] and Social Audit committees).[5]

Jalswarajya is being be implemented in four successive but overlapping phases that follow the pilot phase.[6,7] In phase I, 225 villages were selected from 9 districts. In phases II through IV, an additional 2500 villages are being selected from 28 districts of the total 35 districts in Maharashtra. Thus, the villages that are not selected in phase I but have filed an application provide an important source of pipeline control villages (Chase, 2002). The large number of non-project villages in Maharashtra increases the prospects of finding suitable control villages.

According to the project implementation plan (PIP), there are five stages of project cycle including (i) village selection; (ii) pre-planning and community mobilization, including the establishment of the VWSC; (iii) planning of project activities, appraisal, and sanction; (iv) implementation through procurement, construction and financial management; and (v) operation and maintenance by the VWSC. The overall project cycle for each village is approximately 24 months from selection to implementation.

The villages are selected into Jalswarajya project following their application based on three criteria. First, the village has poor quality and quantity of drinking water and sanitation services. Second, the village has a high proportion of disadvantaged groups (poor and SC/ST populations). Third, the village has sufficient institutional capacity (efficiency in collecting water charges, participation in other community driven projects). In addition to establishing VWSCs and

[5] Other major players include the Water Supply and Sanitation Department and the Reform Support and Project Monitoring Unit (RSPMU) at the state level and the District Water Supply and Sanitation Committee and Zilla Parishads at the district level.

[6] The first phase of the project will be implemented in 9 districts (including the three pilot districts) in approximately 250 villages, a subset of which is the focus of our study. Subsequent phases will include villages from the remaining districts until the target of 2,800 villages are covered by the project.

[7] The project has been piloted since 2002 in 30 villages (10 each from the districts of Thane, Osmanabad, and Satara).

developing village action plans (VAPs), participating villages must meet the cost sharing requirements.

While each village customizes their intervention package, our field visits and discussions with local experts suggest that the chosen interventions will fall into five major intervention clusters because the current WSH situation in all project villages is at the very basic level, all VWSCs are working with a common set of NGOs and project staff, and there is likely to be a natural sequencing of the interventions. Water supply improvements precede other interventions because access to sufficient quantities of flowing water is considered to be a critical input into personal hygiene and sanitation. Collectively, these all point to a narrow and homogeneous set of interventions chosen by project villages. The intervention packages we expect to predominate are:

- An upgrade in water service delivery involving improved availability of water quantity (> 20 lpcd) and hours of supply (> 3 hours / day);
- Improvements in water quality through chlorination, water treatment systems, and regular monitoring of the community source;
- Reduction in open-defecation and greater percentage of community population engaging in fixed point defecation or use of better sanitation facilities through intensive public health education campaign and provision of public facilities;
- Improvements in the basic infrastructure for environmental sanitation through garbage dumps, compost pits, roadside drains and/or soak pits; and
- Information and education about hand washing, hygienic practices for water storage and handling and food preparation, and safe disposal of child feces campaign through school teachers and anganwadi (child care provider in the village) workers.

Our field work also indicates that the upgrade in water service delivery is likely to be combined with one or more of the sanitation, water quality and hygiene education interventions listed above. Thus, each community is expected to adopt a combination of these interventions or an intervention cluster. We will employ a post-stratification plan for identifying 3-4 intervention clusters after the communities implement the WSH interventions. We expect that communities will choose a small set of intervention mixes for exactly the same reasons that they will choose a narrow set of interventions.

3.3. Evaluation Design

Jalswarajya is a typical WSS intervention in that it has multiple components, is community-driven, and will have effects in multiple sectors. Our design draws on advances in quasi-randomized studies and PSM over the last two decades (Rubin 1980, Rosenbaum and Rubin 1983, and Heckman and his colleagues 1979, 1985, 1989, 1997, 1998a, 1998b, 2000, and 2001). As mentioned in Section 2, only a handful of recent evaluations of water and sanitation services have applied a subset of these methods to measure health impacts in rural India and Latin America (Jalan and Ravallion, 2003; Galiani *et al*., 2005; Newman *et al*., 2002). This evaluation study combines PSM, repeated measures, and DID estimation. The research design, which was reviewed by a multidisciplinary advisory team of experts in epidemiology, economics, water and sanitation engineering, and anthropology, also relies on short recall period, analysis of intervention clusters, large sample size, and multi-level control variables.

The evaluation design is guided by a program theory, which is illustrated by the logic model presented in Figure 2. A logic model is a graphical tool for presenting the components of a program and how they are linked to one another (Millar et al. 2001). A logic model is essentially a flow chart or diagram that illustrates relationships between program inputs, outputs, outcomes and impacts.

The logic model in Figure 2 depicts the model assumptions and the linkages between Jalswarajya elements such as WSH interventions and the hypothesized impacts including child health and income. It also helps identify several features of the Jalswarajya project that are relevant to the study such as the goals (e.g., improved livelihoods), target populations (e.g., youth), resources (e.g., multi-lateral loans and governmental institutions), activities (e.g., training and education), outputs (e.g., water and sanitation infrastructure and health and hygiene training), outcomes (e.g., access to water and sanitation infrastructure and hand-washing behaviors), and impacts (e.g., diarrhea rates). In addition, any mediating (e.g., environmental) and intervening (e.g., WSH initiatives by others) factors that could disrupt the flow of this program are identified.

The logic model is useful for identifying indicators for all program components. For example, outcome indicators, which are measured with the household survey, include the distance to water supply and sanitation facilities, and the number of households washing hands, among other things. Impact indicators include the number of children with diarrhea cases and the amount of money households spend on water.

3.3.1 Propensity Score Matching

Matching represents a credible[8] non-experimental method for identifying comparison groups, particularly when randomized designs are not possible because of ethical, political acceptability or logistical reasons. Two features of the Jalswarajya project necessitates non-experimental methods and also introduce selection bias: (1) that communities apply to participate in planning interventions; and (2) that administrators choose communities to implement interventions only if they meet predetermined criteria. The threat of selection bias is introduced because the decision to participate in the intervention is determined in part by the characteristics of the village. The selection bias may be due to observables (or measured characteristics, Z_j, such as age or gender) and/or unobservables[9] (or unmeasured characteristics, V_j, such as motivation or innate ability) (Heckman and Hotz 1989).

PSM accounts for selection based on observables. For every treatment village, a matching village is found from among the non-treatment group based on similarity among a vector of village characteristics. This process accounts for many changes affecting child health in addition to the intervention (e.g., better or worse rainfall, income growth or decline, and the introduction of

[8] Matching has been used in several recent 'good practice' evaluations identified by Judy Baker (2000) and as well as evaluations of various Social Investment Funds with water and sanitation components.

[9] The term "unobservables" is used by Heckman and Hotz (1989) to describe any characteristic that analysts cannot explicitly control for through some measured variable or proxy. It does not necessarily imply selection on a latent construct unless that construct remains unmeasured.

other programs, such as opening or staffing health posts) and assumes that there are no other significant differences between intervention and control units or that any such differences are observable and therefore accounted for.

3.3.2 Difference-in-difference estimation

DID is central to correcting for selection on unobserved characteristics (e.g. residual confounding). It rests on the assumption that the intervention (*i.e*., process of selection into intervention group) and outcomes (*i.e*., impact of intervention on outcome) equations both share a common unobservable component that causes a correlation between the two. The pre-treatment/post-treatment design provides the opportunity to difference the outcome variable and control for this common unobservable term, as well as subtract out any baseline differences such as diarrhea prevalence rates and hygiene behaviors. Thus, the 'treatment effect' is estimated by computing the difference-in-difference (DID) estimator, or the mean difference in the change in diarrhea cases across the intervention and control groups (Ravallion, 2001; Heckman *et al*., 1997, 1998a, 1998b).

3.3.3 Other features

In addition to matching and using a DID estimator, we use four other design features to ensure statistical accuracy in this study. First, we ensure a short recall period by collecting data shortly after the intervention is adequately (as determined by an adequacy assessment) implemented. This allows us to counter any potential threats to validity that could be attributed to other activities or naturally occurring changes over time (*e.g*., 'history' or 'maturation' effects) and improve subject recall on questionnaire items. Second, as we discussed earlier, Jalswarajya is likely to introduce a combination of water supply, sanitation and hygiene education for the rural communities. We use multinomial matching methods so that the intervention combinations can be compared to each other and the non-intervention option, thereby enabling an evaluation of cost-effective treatment alternatives. Third, the large sample size (approximately 10,000 rural households) of this study allows us to detect small changes in health outcomes (<30%) associated with up to four intervention packages. Finally, our data are collected and applied to minimize all likely sources of confounding and estimate unbiased measures of intervention impacts. This dataset includes a variety of individual, household, and community level data on biological, socio-economic, environmental, cultural and institutional factors impact child health outcomes.

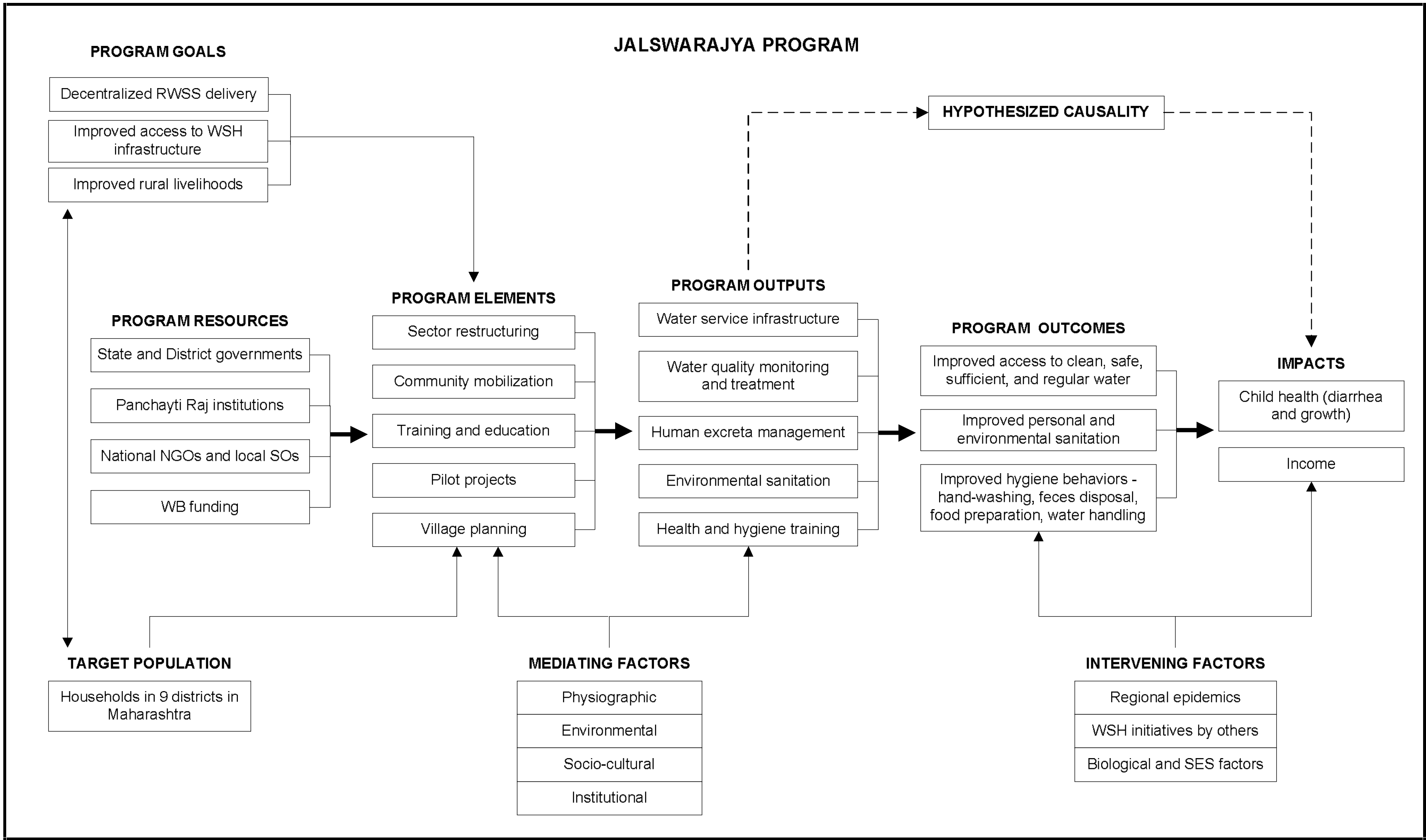
3.4 Sampling and Selection of Control villages

Relying on the best available data on various parameters used in a power calculation, we require an overall sample of approximately 10,000 households for this study. This sample is robust to alternative values of diarrhea prevalence rates and intervention impacts, only as long as we are willing to change the number of intervention packages we study.

3.4.1 Sample Selection

Sample selection involved restriction, matching, and randomization to identify control communities. First, we restricted our sample by excluding coastal and urbanized districts (*e.g.*, Thane, Satara, and Nagpur), which allows us to focus on the rural and dry or drought-prone villages. We selected one district from each of four geographically different regions of Maharashtra. Our restricted sample includes 242 villages from Osmanabad (Marathwada region),

Figure 2. Logic Model for the Quasi-Experimental Impact Evaluation of the Jalswarajya Program



Nashik (Near Mumbai Region), Sangli (Western Maharashtra), and from Buldana (Vidarbha region).

From the restricted sample, we matched each project village with an observationally similar non-project village using propensity score matching. First, we merged data from Jalswarajya MIS and GoI's census of population and of habitats to generate a village level data set comprising 106 variables from 8153 villages (located in 78 blocks and 6 districts) currently in Phase I of the Jalswarajya program.

There are four steps in the PSM based sample selection process. First, estimate a logit model of project participation on a pooled sample of selected project villages and all non-project villages. Second, calculate the propensity score for participating in the Jalswarajya project in the pooled sample. Third, for each project village, find the non-project village with the closest propensity score. Fourth, create a short-list of the non-project villages to serve as the 'matched control' villages.

Our choice of pre-determined variables to be included in the estimation of the model for propensity score estimation relied on the three eligibility criteria for being a Jalswarajya participant: (a) poor quality and quantity of drinking water and sanitation services, (b) high proportion of socially disadvantaged groups, and (c) institutional capacity for fiscal responsibility, community participation and governance. Table 1 lists the 30 variables used in this model. Data on these variables were available for samples of two-sizes: (a) 6200 villages with data on all 30 variables, and (b) 7200 villages with data on all 2001 census variables and water supply coverage. These two data sets are used to construct various matched samples and select the final sample.[10]

Restricting matches to lie within the region of common support (Augurzky and Schmidt, 2001) and trimming 5% of the distribution of estimated propensity scores did not significantly change the set of potential project villages because of the large pool of non-project villages available to provide a match.

Table 2 shows that the estimated logit models are statistically significant, but only explained about 10% of the variation in the data (i.e., pseudo-Rsquared of 0.1). The pseudo-Rsquared was closer to 0.35 for the sub-sample of matched project and control villages. The statistically significant variables had the appropriate signs – that is, they are consistent with the criteria that the Jalswarajya program is supposed to use to select project villages.

We apply a conservative strategy in picking the final list of sample villages by over-sampling control villages. Our final sample of 242 villages comprised of 95 project villages (2 pilot villages, 93 batch 1 villages) and 147 control villages.

[10] We account for the fact that many of the program and administrative decisions are made at the district level through two tasks. First, we include a district dummy variable in the estimation of propensity scores to account for all remaining characteristics unique to the district. Second, we consider two restrictions to the matching process – (a) match within the district, and (b) unrestricted match with any other village in the pool.

The matching strategy is evaluated by checking the balance in covariates (the variables in Table 1) across the project and control villages. We check to see if the matching process reduced (a) the bias (or difference) averaged across Table 1 variables, (b) the bias for key individual variables such as those representing water supply conditions and proportion of socially disadvantaged populations. Table 2 confirms that these criteria were satisfied for all samples retained for further evaluation. Matching reduced bias between 7 and 750 percent, depending on the variable. With one exception, matching eliminated statistically significant differences in the means between treatment and control villages.

3.4.2 Listing and Mapping

We listed and mapped all households in each of the 242 sample villages and identified eligible households– i.e., households with at least one child under five years old – because no such sample frame exists. We interviewed all eligible households in a village if the list was less than 50 and interviewed a random sub-sample of 50 households if the list of eligible households exceeded 50. If a household was not available for an interview then it was visited the same day or next day at different times for up to 3 follow-up visits. If a household was not found or if an interview was refused, then the next household was selected per a pre-established rule of replacement.

3.5 Data Collection Instruments

Household, community, and water quality data were collected via three separate data collection instruments. Four sources of information were used to develop the preliminary household survey instrument – (1) existing survey instruments in the public domain (e.g., LSMS survey, Hygiene Improvement Household Questionnaire, DHS (NFHS in India) survey, India Census), (2) our previous surveys in Nepal and Sri Lanka (Pattanayak et al. 2001; Pattanayak et al. 2004), (3) reviews of various studies on water, sanitation, and hygiene, and (4) inputs from technical advisors and fellow researchers. This preliminary survey instrument was revised using focus group discussions (FGDs), pretests, and feedback during enumerator training.

The final instrument consists of 8 modules.[11] In addition to the household questionnaire, we also designed and implemented a community survey and water quality survey. For the community survey, several questions asked at the household level were modified to be asked at the village level. The community questionnaire sought information on land use, village socioeconomics, water sources, sanitation facilities, social relationships and village institutions. Additionally, we sought detailed and specific information on WSH related projects in the village and general information on any other community based projects through a one page water quality questionnaire to link water samples to individual households and village sources, and to collect sample information including location, cleanliness, and storage.

[11] The 8 household survey modules include: (1) Introduction, Policy Priorities, and Health Literacy; (2) Family Roster and Socio-Demographics; (3) Recent Illness; (4) Water Sources – Quality and Quantity; (5) Water Treatment, Storage, and Hygiene Practices; (6) Sanitation and Sewerage; (7) Socio-Economic Profile; and (8) Community Participation and Preferences.

3.6 Field logistics/Survey implementation

The first round of surveys was conducted between May 4 and June 3, 2005. A Senior Field Manager was in charge of all field operations and development of field routing plans, and was supported by four Field Executives who directly supervised enumerator teams. An enumerator team consisted of 7 individuals to conduct household surveys, 1 supervisor who also conducted the community survey, and 1 individual to collect water samples and complete the water quality survey. In total, 210 people were involved in survey implementation activities.

Completed household and community questionnaires were entered into a Microsoft FoxPro-based data entry template. Validity and accuracy were ensured using tools such as spot checks, range checks, consistency checks, and skip pattern checks. The field offices sent water samples to a laboratory within 24 hours of collection to test for coliform and fecal coliform counts (CFU/100 ml).

4. Conclusions

There are few rigorous scientific impact evaluations showing that WSS policies are effective in delivering many of the desired outcomes. We have considered three reasons for this, all of which are related to the structure of WSS policies and used a case study of a quasi-experimental evaluation of a public WSS intervention in Maharashtra, India to illustrate how quasi-experimental research tools can be applied to address these evaluations.

First, WSS policies are complex. They have multiple objectives, use inputs from multiple sectors, provide a variety of services (water supply, water quality, sanitation, sewerage, and hygiene) using a variety of types of delivery (public intervention, private interventions, public-private partnerships, decentralized delivery, expansion or rehabilitation), and generate effects in multiple sectors (water, environment, health, labor). The case study illustrates the use of careful sample design to enable measurement of the relative contributions of multiple project components (e.g., water supply, sanitation, information) to multiple project effects. Whereas randomizing such a complex mix of project components would be difficult to implement and monitor (and impossible when communities tailor project components), quasi-experimental designs can be designed around multi-sectoral, multi-dimensional projects.

Second, the combination of self-selection by communities and targeting by program administration increases the threat of self-selection bias. This implies that control groups may be difficult to identify and the characteristics of intervention groups are correlated with both their exposure to the intervention and the outcomes that are realized. Quasi-experimental designs that employ PSM and DID can minimize selection bias. The case study reported in the paper shows how pre-intervention PSM also improves the quality of the controls and baselines.

Third, the breadth of effects of WSS policies, which range from greater efficiency in the sector at the national level to improved health at the individual level, implies that data collection strategies and the selection of indicators much be done carefully. The case study shows how the collection of pre-intervention and post-intervention data on communities, households, and individuals, as well as program implementation using a variety of qualitative and quantitative methods is used to triangulate our measures of impacts and controls. This permits a thorough understanding of the program as well as measurement of a broad range of impacts.

We are not advocating that every WSS program and project be evaluated with an experimental or quasi-experimental design, or that every project collect data on outcomes and covariates from treatment and control units before and after the intervention. Our purpose in this paper is to show how tools can be used effectively. There are simply too few impact evaluations in WSS, which impedes our ability to identify, design and justify effective interventions. Each project that builds in the methods and measurements reviewed in the paper will make a small but vital contribution towards filling the large gap in our knowledge about the effectiveness of WSS investments.

5. References

Augurzky, B. and C.M. Schmidt. 2001. The Propensity Score: A Means to an End. IZA Discussion Paper No. 271, Bonn.

Baker, J.L. 2000. Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners. The World Bank, Washington.

Baltazar, J.C. and F.S. Solon. 1989. Disposal of Faeces of Children under Two Years Old and Diarrhoea Incidence: A Case-Control Study. *International Journal of Epidemiology* 18(4 Suppl 2): S16–19.

Blum, D. and R.G. Feachem. 1983. Measuring the Impact of Water Supply and Sanitation Investments on Diarrhoeal Diseases: Problems of Methodology. *International Journal of Epidemiology* 12 (3): 357-365.

Bonilla-Chacin, M., and J. Hammer. 1999. Life and Death among the Poorest The World Bank, mimeo. Washington, DC.

Bosch, C., K. Hommann, G. Rubio, C. Sadoff, and L. Travers. 2000. Water and Sanitation. Chapter 23 in A Sourcebook for Poverty Reduction Strategies Volume 2, pages 371-404. Washington, D.C.: World Bank.

Buse, K. and A. Waxman. 2001. Public-private partnerships: a strategy for WHO. *Bulletin of the World Health Organization* 79(8): 748-754.

Chase, R. 2002. Supporting Communities in Transition: The Impact of the Armenian Social Investment Fund. *World Bank Economic Review* 16(2): 219-240.

Clarke, G., K. Kosec, and S. Wallsten. 2004. Has Private Participation in Water and Sewerage Improved Coverage? Empirical Evidence from Latin America. World Bank Policy Research Working Paper 3445.

Clasen, T., I. Roberts, T. Rabie, W. Schmidt, and S. Cairncross. 2006. Interventions to improve water quality for preventing diarrhea. *The Cochrane Database of Systematic Reviews.* Issue 3, 2006.

Clemens, J.D., and B.F. Stanton. 1987. An Educational Intervention for Altering Water-Sanitation Behaviors to Reduce Childhood Diarrhea in Urban Bangladesh. I. Application of the Case-Control Method for Development of an Intervention. *American Journal of Epidemiology* 125(2):284-291.

Corey, C.G., J.-C. Yang, and S.K. Pattanayak. 2007. Seasonal Variation in Risk Factors Associated with Diarrheal Diseases, Rural Maharashtra, India. RTI Working paper 07-01, Research Triangle Institute, RTP, NC.

Cousens, S.N., T.E. Mertens, and M.A. Fernando. 1990. The Anthropometric Status of Children in Kurunegala District in Sri Lanka: Its Relation to Water Supply, Sanitation and Hygiene Practice. *Tropical Medicine and Parasitology* 41(1): 105–114.

Curtis, V., and S. Cairncross. 2003. Effect of washing hands with soap on diarrhea risk in the community, a systematic review. *Lancet Infectious Disease.* 3:275-281.

Daniels, D.L., S.N. Cousens, L.N. Makoae, and R.G. Feachem. 1990. A Case-Control Study of the Impact of Improved Sanitation on Diarrhoea Morbidity in Lesotho. *Bulletin of World Health Organization* 68(4): 455–463.

Dasgupta, P. 2004. A Case-Control Study of the Impact of Improved Sanitation on Diarrhoea Morbidity in Lesotho. *Environment and Development Economics* 9: 83-106.

Dickinson, K.L., S.K. Pattanayak, J.-C. Yang, S.R. Patil, P. Praharaj, R. Mallick, J.L. Blitstein, and C. Poulos. 2007. Nature's Call: Can Social Mobilization Promote Toilet Use and Reduce Child Diarrhea? Results from a Community Randomized Trial in Orissa, India. RTI Working paper 07-, Research Triangle Institute, RTP, NC.

Dickinson, K.L., S.K. Pattanayak, P. Praharaj, R. Mallick, and S.R. Patil. 2006. Community-led Total Sanitation in Bhadrak, Orissa: Multiple Perspectives on the Nature of the Intervention and Its Initial Results. RTI Working paper 06-, Research Triangle Institute, RTP, NC.

Drees, F., J. Schwartz, and A. Bakalian. 2004. Output-based Aid in Water: Lessons in Implementation from a Pilot in Paraguay. Viewpoint No. 270. Washington, D.C.: World Bank.

Esrey, S. 1996. Water, Waste and Well-being: A Multi-Country Study. *American Journal of Epidemiology* 143(6):608-623.

Esrey, S.A., J.B. Potash, L. Roberts et al. 1990. Health BeneFits from Improvements in Water Supply and Sanitation: Survey and Analysis of the Literature on Selected Diseases, WASH Technical Report No. 66, Washington DC, Water and Sanitation for Health Project.

Esrey, S.A. and J.-P. Habicht. 1986. Epidemiologic evidence for health benefits from improved water and sanitation in developing countries. *Epidemiologic Reviews* 8, 117–128.

Esrey, S., J. Potash, L. Roberts, and C. Shiff. 1991. Effects of Improved Water Supply and Sanitation on Ascariasis, Diarrhea, Dracunculiasis, Hookworm Infection, Schistosomiasis, and Trachoma. *Bulletin of the World Health Organization* 69(5): 609-621.

Ferraro, P.J., and S. K. Pattanayak. 2006. Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments. *PLOS Biology* 4(4): e105 (0482-0488).

Fewtrell L. and J.M. Colford. 2004. Water, Sanitation and Hygiene: Interventions and Diarrhoea. Health, Nutrition and Population Discussion Paper. <http://www1.worldbank.org/hnp/Pubs_Discussion/Fewtrell&ColfordJuly2004.pdf>.

Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L and Colford, J. M. 2005. Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *The Lancet infectious diseases* 5: 42-52.

Galiani, S., P. Gertler, and E. Schargrodsky. 2005. Water for Life: The Impact of the Privatization of Water Services on Child Mortality. *Journal of Political Economy* 113: 83-120.

Government of India. 2002. Tenth Five Year Plan (2002-2007). Planning Commission, Government of India. <http://planningcommission.nic.in/plans/planrel/fiveyr/10th/default.htm>.

Han, A.M. and K. Moe. 1990. Household Faecal Contamination and Diarrhoea Risk. *Journal of Tropical Medicine and Hygiene* 93: 333–336.

Heckman, J. 1979. Sample Selection Bias As A Specification Error. *Econometrica* 47(1):153–161.

Heckman, J. 2000. Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective. *Quarterly Journal of Economics* 115:45–97.

Heckman, J., and J. Hotz. 1989. Choosing among Alternative Methods of Estimating the Impact of Social Programs: The Case of Manpower Training. *Journal of the American Statistical Association* 4(408):862–874.

Heckman, J., and R. Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions: An Overview." *Journal of Econometrics* 30(1-2):239–267.

Heckman, J., H. Ichimura, and P. Todd. 1997. Matching As An Econometric Evaluation Estimator: Evidence from Evaluating A Job Training Program. *Review of Economic Studies* 64(4):605–654.

Heckman, J., H. Ichimura, and P. Todd. 1998b. Matching As An Econometric Evaluation Estimator. *Review of Economic Studies* 65:261–294.

Heckman, J., H. Ichimura, J. Smith, and P. Todd. 1998a. Characterizing Selection Bias Using Experimental Data. *Econometrica* 66(5):1017–1098.

Heckman, J., J. Tobias, and E. Vytlacil. 2001. Four Parameters of Interest in the Evaluation of Social Programs. *Southern Economic Journal* 68(2):211–223.

Hughes, G., K. Lvovsky, and M. Dunleavy. 2001. Environment Health In India: Priorities in Andhra Pradesh. South Asia Environment and Social Development Unit, The World Bank.

Huttly, S.R., S. Morris, and V. Pisani. 1997. Prevention of Diarrhea in Young Children in Developing Countries. *Bulletin of the World Health Organization* 75(2): 163–174.

International Institute for Population Sciences (IIPS) and ORC Macro. 2000. National Family Health Survey (NFHS-2), 1998-1999. IIPS, Mumbai, India.

Jalan, J. and E. Somanathan. 2004. The importance of being informed. Experimental evidence on the demand for environmental quality. SANDEE Working Paper 04-08.

Jalan, J. and M. Ravallion. 2003. Does Piped Water Reduce Diarrhea for Children in Rural India? *Journal of Econometrics* Vol. 112(1): 153-173.

Jensen, P.K. et al. 2002. Domestic Transmission Routes of Pathogens: The Problem of In-House Contamination of Drinking Water During Storage in Developing Countries. *Tropical Medicine and International Health* 7(7): 604-9, July 2002.

Kaufmann, R. 2005. Water, Sanitation and Hygiene Interventions for Health – What Works? Environment Matters. The World Bank Group. Annual Review. July 2004-June 2005: 24-25.

Lechner, M., F. Pfeiffer, H. Spengler, & M. Almus. 2000. The Impact of Non-Profit Temping Agencies on Individual Labour Market Success in the West German State of Rhineland-Palatinate. Papers 00-02, Centre for European Economic Research.

Mansuri, G., and V. Rao. 2004. Community–Based and –Driven Development: A Critical Review. *The World Bank Research Observer* 19(1):1–39

McKenzie, D. and I. Ray. 2005. Household Water Delivery Options in Urban and Rural India. Stanford Center for International Development Working Paper no. 224. Stanford University: Stanford, CA.

Millar, A., R. Simeone, and J. Carnevale. 2001. Logic models: a systems tool for performance management. *Evaluation and Program Planning* 24: 73-81.

Mumssen, Y. 2004. Output-based Aid in Cambodia: Private operators and local communities help deliver water to the poor. OBApproaches No. 01. Washington, D.C.: World Bank.

Newman, J., M. Pradhan, L.B. Rawlings, G. Ridder, R. Coa, and J.L. Evia. 2002. An Impact Evaluation of Education, Health, and Water Supply Investments by the Bolivian Social Investment Fund. *World Bank Economic Review* 16: 241–274.

Patil, S.R., and S.K. Pattanayak. 2007. The In-House Water Quality Management: Uncovering the Complex Web of Household Behaviors, and Water and Sanitation Interventions. RTI Working Paper 07-, Research Triangle Institute, RTP, NC.

Patil, S.R., S.K. Pattanayak and S. Vinerkar. 2006. Gauging Adequacy of Community Water Supply and Sanitation Projects in Maharashtra: Methodological Triangulation. RTI Working Paper 06-, Research Triangle Institute, RTP, NC.

Pattanayak, S.K., J.-C. Yang, S. R. Patil, K. M. Jones, C. P. Poulos, C. G. Corey and R. Kwok. 2005a. Environmental health outcomes of community-led improvements in rural water supply, sanitation, and hygiene in Maharashtra and Orissa. Study protocol prepared for the World Bank, South Asia Environment and Social Development Unit.

Pattanayak, S.K., J.-C. Yang, D. Whittington, and K.C. Bal Kumar. 2005b. Coping with Unreliable Public Water Supplies: Averting Expenditures by Households in Kathmandu, Nepal. *Water Resources Research* 41, W02012, doi: 10.1029/2003WR002443.

Pattanayak, S.K., J.L. Blitstein, J. Yang, S.R. Patil, K.M. Jones, C. Poulos, and K.L. Dickinson. 2006. Evaluating Information and Communication Strategies to Promote Latrine Use and Improve Child Health: Design and Baseline Findings from a Community Randomized Trial in Bhadrak, Orissa. RTI Working Paper 06-05, Research Triangle Institute, RTP, NC.

Poulos, C., S.K. Pattanayak, and K. Jones. 2006. A Guide to Water and Sanitation Sector Impact Evaluations. Doing Impact Evaluation Series No. 4. World Bank. < http://siteresources.worldbank.org/INTISPMA/Resources/383704-1146752240884/Doing_ie_series_04.pdf>

Prennushi, G., G. Rubio, and K. Subbarao. 2000. Monitoring and Evaluation. Chapter 3 in A Sourcebook for Poverty reduction Strategies Volume 1, pages 105-130. Washington, D.C.: World Bank.

Ravallion, M. 2001. The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation. *World Bank Economic Review* 15(1): 115–140.

Rawlings, L.B., L. Sherburne-Benz, and J.V. Domelen. 2004. Evaluating Social Funds: A Cross-Country Analysis of Community Investments. The World Bank, Washington, D.C.

Reform Support and Project Management Unit (RSPMU). 2004. Maharashtra Rural Water Supply and Sanitation Project – Jalswarajya – Quarterly Progress Report, January – March 2004. RSMPU, Water Supply and Sanitation Department, Government of Maharashtra.

Rosenbaum, P.R. and D.B. Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70(1): 41-55.

Rubin, D.B. 1980. Bias Reduction Using Mahalanobis-Metric Matching. *Biometrics* 36: 293–298.

Sobsey, M.D. 2002. Managing Water in the Home: Accelerated Health Gains from Improved Water Supply. WHO/SDE/WSH/02.02, World Health Organization, Geneva.

The United Nations (UN) 2005. The millennium Development Goals Report. UN: New York. Accessed at http://unstats.un.org/unsd/mi/pdf/MDG%20Book.pdf on January 25, 2006.

The United Nations Development Programme (UNDP) 2006. Human Development Report 2006 – Beyond scarcity: Power, poverty and the global water crisis. UNDP: New York. Accessed at < http://hdr.undp.org/hdr2006/report.cfm> on 31 January 2007.

The World Bank – Operations Evaluation Department. 2004. Monitoring and Evaluation: Some tools, methods, and approaches. The World Bank, Washington, D.C.

Van der Klaauw, B. and L. Wang. 2003. Child Mortality in Rural India. Working paper. Free University Amsterdam and World Bank.

VanDerslice J and Briscoe J. 1993. All coliforms are not created equal: A comparison of the effects of water source and in-house water contamination on infantile diarrheal disease. *Water Resources Research*, 29(7):1983-95.

Wang, L. 2002. Health Outcomes in Low-income Countries and Policy Implications: Empirical Findings from Demographic and Health Surveys. Policy Research Working Paper #2831, The World Bank.

Water and Sanitation Program (WSP). 2003. Public-Private Partnerships for Health: A Review of Best Pracices in the Health Sector.

Young, B. and J. Briscoe. 1987. A Case-Control Study of the Effect of Environmental Sanitation on Diarrhoea Morbidity in Malawi. *Journal of Epidemiology and Community Health* 42(1): 83–88.

Table 1. Variables Used in the Propensity Score Estimation

Description	Logit Model using 2001 Census, 1991 Census, and 1999 Habitat Data	Logit Model using 2001 Census
Dependent Variable	Dummy variable indicating whether village is in Jalswarajya Project	
Proportion of males in the village in 2001	-8.24 (4.36)*	-5.33 (3.97)
Proportion of children (less than 6 years) in the village in 2001	1.56 (3.29)	0.75 (3.22)
Proportion of scheduled castes in the village in 2001	1.58 (0.80)**	1.56 (0.74)**
Proportion of scheduled tribes in the village in 2001	1.99 (0.44)***	1.91 (0.41)***
Proportion of female workers in the village in 2001	-0.19 (1.59)	-0.64 (1.44)
Proportion of cultivators in the village in 2001	-0.71 (0.74)	-0.84 (0.68)
Proportion of agricultural labors in the village in 2001	0.82 (0.78)	0.82 (0.71)
Proportion of marginal workers in the village in 2001	0.71 (0.58)	0.53 (0.54)
Number of households in the village in 2001	-0.00 (0.00)	0.00 (0.00)
Average household size in the village in 2001	0.35 (0.16)**	0.31 (0.15)**
Female literacy rate in the village in 2001	0.01 (0.01)	0.00 (0.01)
Proportion of permanent houses at the block level in 2001	-0.87 (0.82)	-1.01 (0.77)
Proportion of households with private tap at the block level in 2001	0.00 (1.06)	-0.67 (0.99)
Proportion of households without latrine at the block level in 2001	-5.57 (1.83)***	-3.71 (1.69)**
Proportion of households with electricity at the block level in 2001	-1.27 (1.46)	0.13 (1.29)
Proportion of households who use firewood / crop residue / cowdung cake as cooking fuel at the block level in 2001	-1.41 (1.58)	-1.39 (1.45)
Water supply level (lpcd) in the village	-0.05 (0.01)***	-0.04 (0.01)***
Distance from nearest water source in the village	0.00 (0.03)	
Water affected by Arsenic in the village	0.86 (0.73)	
Water affected by Fluorides in the village	0.65 (0.69)	
Water affected by Nitrate in the village	1.04 (0.73)	
Water affected by salinity in the village	-0.43	

Description	**Logit Model using 2001 Census, 1991 Census, and 1999 Habitat Data**	**Logit Model using 2001 Census**
Dependent Variable	Dummy variable indicating whether village is in Jalswarajya Project	
	(1.04)	
Water affected by odor in the village	0.89 (0.62)	
Availability of market facility in the village in 1991	-1.02 (0.22)***	
Availability of bus stop in the village in 1991	-0.11 (0.24)	
Availability of railway station in the village in 1991	-1.27 (1.04)	
Presence of pucca roads in the village in 1991	0.24 (0.21)	
Presence of kachcha roads in the village in 1991	-0.01 (0.01)*	
Area of the village in 1991	0.00 (0.00)	
Village in Buldhana District	0.72 (0.38)*	0.82 (0.33)***
Village in Nashik District	-0.08 (0.43)	-0.46 (0.41)
Village in Osmanabad District	0.87 (0.50)*	0.64 (0.47)
Village in Sangli District	1.13 (0.41)***	1.12 (0.36)***
Constant term	7.19 (3.96)*	2.33 (3.61)
Number of observations	6201	7181
Pseudo R2	0.1091	0.0799

Table 2. Covariate Balancing using Pooled Model

Variable	Sample	Mean		%bias	%reduct |bias|	t-test	
		Treated	Control			t	p>|t
Proportion of males in the village in 2001	Unmatched	.50837	.50731	3.9		0.29	0.774
	Matched	.50837	.50752	3.1	20.2	0.29	0.769
Proportion of children (<6 years) in the village in 2001	Unmatched	.16621	.1529	33.8		3.50	0.000***
	Matched	.16621	.16189	11.0	67.5	0.74	0.463
Proportion of scheduled castes in the village in 2001	Unmatched	.1036	.1008	2.7		0.23	0.821
	Matched	.1036	.0798	22.8	-748.8	1.78	0.077
Proportion of scheduled tribes in the village in 2001	Unmatched	.28646	.18353	29.4		3.15	0.002**
	Matched	.28646	.30814	-6.2	78.9	-0.36	0.717
Proportion of female workers in the vill. in 2001	Unmatched	.45324	.45597	-3.9		-0.32	0.752
	Matched	.45324	.45372	-0.7	82.5	-0.05	0.963
Proportion of cultivators in the village in 2001	Unmatched	.48628	.52535	-18.5		-1.65	0.099*
	Matched	.48628	.52176	-16.8	9.2	-1.18	0.241
Proportion of agricultural labors in the village in 2001	Unmatched	.39065	.31317	35.6		3.51	0.000***
	Matched	.39065	.34188	22.4	37.1	1.55	0.124
Proportion of marginal workers in the vill. in 2001	Unmatched	.16556	.18379	-11.3		-0.98	0.329
	Matched	.16556	.17479	-5.7	49.4	-0.37	0.709
Number of households in the village in 2001	Unmatched	384.87	291.42	26.1		2.46	0.014***
	Matched	384.87	387.12	-0.6	97.6	-0.03	0.974
Average household size in the village in 2001	Unmatched	5.3215	5.0309	44.1		3.94	0.000***
	Matched	5.3215	5.3188	0.4	99.1	0.03	0.977
Female literacy rate in the village in 2001	Unmatched	51.924	57.267	-33.7		-3.55	0.000**
	Matched	51.924	54.311	-15.1	55.3	-1.05	0.296
Proportion of permanent houses at the block level in 2001	Unmatched	.39369	.42923	-19.6		-1.86	0.064*
	Matched	.39369	.40921	-8.6	56.3	-0.57	0.569
Proportion of households with private tap at the block level in 2001	Unmatched	.41704	.50246	-51.9		-4.82	0.000***
	Matched	.41704	.41533	1.0	98.0	0.07	0.945
Proportion of households without latrine at the block level in 2001	Unmatched	.8295	.81303	22.9		2.07	0.038**
	Matched	.8295	.83512	-7.8	65.9	-0.49	0.627
Proportion of households with electricity at the block level in 2001	Unmatched	.56535	.66577	-57.8		-5.53	0.000***
	Matched	.56535	.54211	13.4	76.9	0.82	0.414
Proportion of households who use firewood / crop residue / cowdung cake as cooking fuel at the block level in 2001	Unmatched	.86151	.82232	37.0		3.39	0.001***
	Matched	.86151	.8556	5.6	84.9	0.40	0.692
Water supply level (lpcd) in the village	Unmatched	28.396	33.301	-36.7		-3.30	0.001***
	Matched	28.396	23.835	34.1	7.0	2.33	0.021**