Can Biometric Tracking Improve Healthcare Provision And Data Quality? Experimental Evidence from Tuberculosis Control in India

Thomas Bossuroy
Vincent Pons
Clara Delavallade

Working Paper 19-077
Can Biometric Tracking Improve Healthcare Provision And Data Quality? Experimental Evidence from Tuberculosis Control in India

Thomas Bossuory
World Bank

Clara Delavallade
World Bank

Vincent Pons
Harvard Business School

Working Paper 19-077
CAN BIOMETRIC TRACKING IMPROVE HEALTHCARE PROVISION AND DATA QUALITY?

EXPERIMENTAL EVIDENCE FROM TUBERCULOSIS CONTROL IN INDIA¹

Thomas BOSSUROY
(World Bank)

Clara DELAVALLADE
(World Bank)

Vincent PONS
(Harvard Business School)

Developing countries increasingly use biometric identification technology to improve the reliability of administrative information and deliver social services more efficiently. This paper exploits the random placement of biometric devices in tuberculosis (TB) treatment centers in urban slums across four Indian states. The devices record health worker attendance and patient adherence to protocol, and they automatically prompt follow-up treatment. We combine data from surveys, independent field visits, and government registers to identify impacts on TB-control outcomes and the quality of data. First, we find that patients enrolled at centers with biometric monitoring are 25 percent less likely to interrupt their treatment than those at regular centers – an improvement driven by increased attendance and efforts by health workers and greater protocol adherence by patients. Second, the biometric devices decrease overreporting of patient numbers in both NGO data and government registers by 20 percent, and underreporting of treatment interruptions by 25 percent. We find no effect on worker or patient satisfaction. This suggests biometric technology is both an effective and sustainable way to improve the state capacity to deliver healthcare in challenging areas.

¹ We are deeply grateful to Sandeep Ahuja, Shelly Batra and Prateek Ahuja at Operation ASHA for initiating and supporting this study. For comments and suggestions, we thank Marcella Aslan, Jishnu Das, Puneet Dewan, Esther Duflo, Rifat Hasan, Michael Kremer, Manoj Mohanan, Patrick Mullen, Madhukar Pai, Dr KS Sachdeva, Bill Thies, as well as audiences at Harvard, Warwick, the World Bank, GREThA, NEUDC, PAA, Duke University and the Hewlett Foundation Workshop on Quality of Healthcare and USAID’s TB Industry Day. For excellent field management and research assistance, we are grateful to Siddhartha Baral, Suthi Chandrasekaran, Rachna Nag Chowduri, Sadish Dhakal, Radhika Jain, Chand Mazumdar, Mansi Shah and Ameeek Singh. Miral Kalyani and Umry Shukla analyzed the qualitative data. Ravi Agarwal carried out scalability analysis. This project was funded by grants from USAID, AusAid and the Health Results Innovations Trust Fund administered by the World Bank. Ethical approvals were sought and obtained from the MIT Committee on the Use of Humans as Experimental Subjects (COUHES), the Institutional Review Boards of the Institute for Financial Management and Research (IFMR) in India and the University of Cape Town, South Africa. This study is registered in the AEA RCT Registry and the unique identifying number is AEARCTR-0000038. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the World Bank or the Harvard Business School. All errors are ours.
1- Introduction

Principle-agent theory posits that employees may reduce effort in the presence of information asymmetry (Alchian and Demsetz, 1972; Laffont and Martimort, 2002; Prendergast, 1999). This not only has damaging consequences for the provision of social services in programs with low monitoring capacity (Chaudhury and Hammer 2004, Banerjee et al. 2004), it also creates incentives for the agent to alter the reporting of performance data (Brian and Levitt 2003; Kleven et al. 2011). Misreporting in the public sector hampers the government’s ability to measure and address the needs of its citizens, arguably a key component of state capacity (Porter 1995; Desrosières 2002).

These issues are especially consequential in tuberculosis-control programs. With 1.3 million deaths per year, tuberculosis (henceforth TB) is the second deadliest communicable disease globally, with India having the highest number of cases. About 1,100 Indians still die of TB every day (WHO, 2018) – as a comparison, the number of TB deaths every ten days in India is roughly equal to the number of Ebola deaths during the entire 2014-2015 Ebola outbreak in West Africa. Yet most cases of TB can be cured with a range of cheap, widely available antibiotics. For this, patients need to be diagnosed and then comply with a six-month treatment regimen, which calls for strong support services at the local level. To accomplish effective TB control, the World Health Organization promotes a strategy called Directly Observed Therapy, Short Course (DOTS) which calls for widely decentralized healthcare systems backed with dedicated record management systems around the world.

India’s Revised National TB Control Program (RNTCP) has state-level and district-level offices across the country. It tasks community-based health workers to refer potential cases to detection centers and observe that patients take their treatment every few days. However, the program’s limited monitoring capacity – compounded with low-quality healthcare in the country (Mohanan et al., 2016; Das and Hammer, 2007) and particularly poor TB care (Satyanarayana et al., 2015; Das et al., 2015) – might stymie the efficiency of TB Control Program in India with direct consequences on morbidity and survival. Failure to detect cases on time and interruption of the treatment, are highly likely to result in the death of the patient and the contamination of several other individuals.

This paper examines the benefits of using a biometric monitoring device to improve service delivery and reduce misreporting in TB control efforts in India. We partnered with Operation ASHA, the largest NGO delivering primary care to TB patients in India, to randomize the roll-out of biometric devices across 131 treatment centers, each covering a population of over 20,000 individuals, located in urban slums across four states in Northern India. The technology was designed to perform three main functions: ensuring that patients themselves received their drugs from the health worker; generating alerts when they failed to take
their pills, which facilitated rapid follow-up by health workers; and creating a real-time tool for program managers to monitor attendance and performance of health workers. The devices also drastically cut the scope for health workers to alter health records by over-reporting new cases or underreporting instances of treatment interruption.

Biometric tracking may improve the quality of care, and reliable identification of patients would be especially important for delivering financial and nutritional support (RNTCP 2017). However, this technology may also have unintended consequences. Patients weakened by the disease may no longer be able to take home several doses at a time for self-administration or send a relative to the center to pick up their pills, which may lead to more strain and more frequent treatment interruptions. Health workers focusing on ensuring treatment adherence may reallocate efforts away from other important activities such as identifying early cases, or they may strategically avoid enrolling patients with a presumed higher likelihood to miss pills and interrupt their treatment.

To sort out these conflicting forces, we use information covering 4911 patients and 85 health workers from the unique combination of a variety of sources: several rounds of survey data collected both on patients and health workers, program data collected from the NGO, administrative data collected from government hospitals, and independent data collected through random spot checks and observation days at DOTS centers. The combination and comparison of program-generated and independently collected data allows us to reliably estimate impacts on both real TB control outcomes and on the extent of misreporting in public health data.

The empirical analysis yields two main findings. First, biometric tracking increases patient adherence to the treatment regimen: patients’ interruption of their TB treatment reduces by 25 percent, and in-person pill intake, a core requirement of the DOTS protocol, improves by about 25 percent. These results appear to be driven by a 20-percent boost in provider attendance at the center and an increased frequency of home visits to patients. Second, biometric tracking measurably decreases misreporting in NGO data and official government registers. The magnitude of the reduced discrepancies between these reported data and our own outcome measurements indicates that, absent the devices, at least 20 percent of all default cases go unreported in government registers and about 20 percent of all new patients declared are overreported.

For both of these main findings we explore the mechanisms driving treatment effects and rule out alternative explanations. As suggested by multitasking models in the agency literature (Holmström and Milgrom, 1991; Van Den Berg and Van Der Klaauw, 2006), we test whether health workers strategically reallocated their efforts towards the tasks monitored by the device, namely center attendance and patient treatment adherence, at the expense of active case finding. However, we do not find any significant negative impact
on the number of patients detected or their observable characteristics. This decreases the likelihood that our impact estimates may be confounded by differential patient selection induced by the intervention, e.g. if patients selected by health workers operating biometric centers were more likely to complete their treatment because they live closer to the DOTS center. To bring further support for this conclusion, we check that our main results are robust to restricting the analysis to the subset of patients who were enrolled prior to the experiment (but who were still undergoing treatment) as, by construction, these patients’ selection could not be affected by the intervention.

Finally, we provide positive evidence on the sustainability and acceptability of the technology. The decrease in reported patient detections reduced the earnings of health workers with biometric devices by 7 percent. But the devices also reduced their workload so that, overall, their job satisfaction did not drop. Patients did not report lower levels of satisfaction with their treatment either, despite having to visit the treatment center themselves more frequently. Consistent with maintained health worker and patient satisfaction, the technology remained in place and functional over time and its impact did not taper off significantly over the duration of the experiment.

The main contribution of this paper relates to the literature on public administration and state capacity in developing countries, which is receiving renewed attention. With millions left out of growth-induced poverty reduction trends, social programs responsive to the needs of the poor are becoming of critical importance to eradicate poverty and build human capital. As the primary provider or regulator of social infrastructure, the state is called to play a critical role that requires increased capacity and accountability (Besley and Persson 2010; Page and Pande 2018). However, the capacity of public-sector agencies to efficiently deliver services to the poor and vulnerable is a challenge (World Bank 2004) which results in vast inequalities in access to healthcare across countries and along the wealth distribution (World Bank 2017).

This paper addresses a first critical dimension of state capacity, namely the efficiency of administrative structures in developing countries and their capacity to serve the poor. We offer evidence that a simple intervention can improve the alignment of individual behaviors and institutional incentive structures – key elements of state capacity (Finan et al. 2015, Rasul and Rogger 2016) – to better deliver services for vulnerable populations. Our results add to the literature that shows how system reforms, especially when facilitated by technology, can achieve measurable progress for service delivery by the public sector (Muralidharan et al. 2014, Duflo et al. 2013, Banerjee et al. 2015, Banerjee et al. 2016).

More specifically, this paper contributes to the literature on absenteeism in social sectors (Banerjee and Duflo 2006), a key factor undermining state capacity to deliver services on which impact evaluations have
shown diverging results. On the one hand, technology-based attendance monitoring has shown positive impacts on the presence of teachers (Duflo, Hanna and Ryan, 2012) and health workers (Banerjee, Duflo and Glennerster, 2008, Dhaliwal and Hanna, 2017, Callen et al. 2016). These results stand in contrast to the very mixed evidence coming from attempts to strengthen service delivery through community-based monitoring (Bjorkman and Svensson 2009, Rafter et al. 2019, Banerjee, Deaton, and Duflo, 2004, Kremer and Vermeersch, 2005). Both sets of interventions emphasize that monitoring works better in the presence of other financial incentives. Our paper extends this literature by exploring if positive results can be achieved by combining monitoring with productivity gains for the health worker, instead of explicit financial gains.

The results presented here further add to the evidence on ways to boost medication adherence, a consequential challenge directly linked to state capacity (Sabate 2013) that receives much policy attention but for which “even the most effective interventions have modest effects” (McDonald et al. 2002). Demand-side interventions, targeted to the patients, include the provision of tailored information on pill packaging (Cohen and Saran 2018), text message reminders (Raifman et al., 2014; Pop-Eleches et al., 2011), or financial incentives (Giuffrida and Torgerson, 1997; DeFulio and Silverman, 2012). This study contributes to the less-studied field of supply-side interventions targeting allied healthcare providers, which are generally part of larger intervention packages and whose impacts are less well documented (Nieuwlaat et al. 2014).

We also contribute to a larger literature on leveraging technology to increase the state capacity to deliver services. Two aspects of technology are used in the intervention evaluated here: connectivity that enables closer oversight of service delivery in remote areas and biometric identification that enables the creation of reliable registers of assistance beneficiaries and their adherence to treatment. In the health sector, mobile technology has been used to send text message reminders to patients, which helped boost adherence to antimalarial treatment in Northern Ghana (Raifman et al., 2014) and antiretroviral treatment in Kenya (Lester et al., 2010; Pop-Eleches et al., 2011). Mobile communication can also be targeted to health workers, as was the case in Ethiopia with automated messages to remind health workers of key appointments with pregnant women and newborns and help track the stock of essential medicines (Otto et al., 2015), or in Kenya with messages to improve health worker’s adherence to guidelines for malaria treatment (Zurovac et al., 2011). Biometric identification has seen a rapid growth in the past decade and has now been used as a mechanism for delivering public programs in more than 80 developing countries, with applications ranging from financial services to social transfers, civil service reforms, and health policies (Zuniga et al., 2010; Gelb and Clark, 2013). Evidence on the impact of biometric identification for beneficiaries is encouraging but still scarce. More specifically, the use of biometric devices for TB control was studied...
through qualitative beneficiary surveys (Bhatnagar et al. 2012) and observational studies (Snidal et al. 2015). This paper innovates in using a randomized design to solve selection issues and rigorously quantify the impact of the technology. Outside of the health sector, studies have shown biometric technology to increase the efficiency of payments in cash-for-work and pension programs (Muralidharan et al., 2014) and fuel subsidies (Barnwal, 2015) in India, and improve credit repayment rates in Malawi (Gine et al., 2012).

A second, less studied, dimension of state capacity pertains to the ability to measure citizens’ needs and public-sector performance, and base policy decisions on a sound assessment of priorities. Historians of developed countries have shown how the construction of public statistical systems was a driving force behind the establishment of modern states (Tooze 2001, Desrosières 2002). In the context of health systems in India, Morton et al. (2016) show how data from a hospital insurance program are practically unusable even though simple fixes could turn them into critical information for state efforts to improve healthcare for the poor. By significantly strengthening the quality and reliability of government health registers, the intervention we study may improve state capacity in the consequential area of TB control.

By measuring impacts on the accuracy of administrative data, we also advance the literature on misreporting and cheating. Becker and Stigler (1974) proposed a seminal model linking fraudulent behavior both to the nature of monetary incentives attached to corruption and to the likelihood and consequences of being caught. Among the related literature, our paper is closest to Olken (2007), Duflo et al. (2013) and Banerjee et al. (2016) which study fraudulent misreporting in the context of road construction in Indonesia, industrial plant pollution in Gujarat, and fiscal transfer systems in Bihar, respectively. Our study yields the consistent conclusion that a stronger top-down monitoring of adherence, be it through strengthened audit processes or strengthened reporting and accounting mechanisms, improves the reliability of measurement but also the performance and efficiency of public delivery systems. Our measure of misreporting used in this paper is also essentially similar to the measures used in these three papers, in that it is based on the comparison between official government data and independently collected data for the same outcomes of interest. A comparable measurement strategy was also used to identify the manipulation of GDP growth statistics by comparing it to nighttime luminosity data (Henderson 2012, Martinez 2019). Other relevant findings on monitoring and cheating come from public hospitals in Argentina (Di Tella and Schargrodsky, 2003), voter registration in Ghana (Ichino and Schündeln 2012), tax evasion in the U.S. (Slemrod et al. 2001) and in Denmark (Kleven et al. 2011), telephone call centers in the U.S. (Nagin et al. 2002), and electoral fraud in Afghanistan (Callen and Long 2015). The impact of monetary incentives to misreport performance have been studied in the context of firm financial restatements (Burns and Kedia, 2006), teacher cheating (Jacob and Levitt, 2003) and in a lab experiment (Maas and Van Rinsum, 2013).
From a policy perspective, our results contribute to ongoing efforts in India to strengthen the delivery of social programs through biometrically enhanced mechanisms. They point to the benefits that unique identification such as India’s Aadhaar system may have on TB control programs, both to facilitate the delivery of primary care and to enhance dramatically the reliability of the national register of TB patients. This latter point is of even greater importance as direct benefits transfers to TB patients will be implemented to incentivize adherence (RNTCP 2017) and could also be used to encourage patient referrals (Goldberg et al. 2018). Policy implications of our study for TB control programs in India are discussed in more detail in Section 6.

The remainder of the paper is organized as follows. Section 2 provides background on tuberculosis in India and on the intervention evaluated in this paper. Section 3 describes the experimental setup, measurement strategy and internal validity checks. Section 4 presents the results on TB control outcomes and provider efforts. Section 5 discusses the empirical strategy and findings related to misreporting. Section 6 presents evidence on the acceptability, sustainability and scalability of the intervention, and Section 7 concludes.

2- Setting: policy challenge and intervention

a. TB control in India

TB remains the largest infectious killer of adults, claiming nearly 1.3 million lives annually, of which about 410,000 are in India (WHO, 2018). While severe coughing is the hallmark of the disease, those infected also lose weight, become listless and endure crippling pain. Over 80 percent of the patients we surveyed in this study reported suffering from respiratory problems, body pain, and other general health problems. As a result, affected people are often unable to work or care for their families, with direct consequences for household livelihoods. About half of the patients who were gainfully employed at the time they contracted TB reported losing more than 80 percent of their workdays and income as a direct consequence of the disease.

Roughly 97 percent of all new TB cases are of drug-susceptible strains of the disease, and effective first-line antibiotics are available at no cost in India. However, two major challenges hinder faster progress in the fight against TB: detecting infection and enrolling patients in treatment early in the disease’s progress, and ensuring that they complete the full course of treatment. Efficient public health programs are necessary to meet both of these challenges.

The first challenge, known as early case finding, is critical to improving the condition of the infected patient, but also to prevent the spread of the highly contagious, airborne disease. Contagion is particularly rapid in
urban slums characterized by high population density and poor sanitation. In our study sample, the average patient lives in a two-room house with five other family members. Fewer than a quarter of patients have in-home access to tap water. As a result, disease transmission takes places mostly within families and communities. Over 40 percent of the TB patients we surveyed reported that at least one other member of their family had a history of the disease.

Early case finding is hampered by low levels of awareness and information on TB, poor access to care and preventive services, and fear of social stigma about contracting the disease. Only about half of the patients know about the most prominent symptoms of TB before starting its treatment, and only two-thirds know that it is a communicable disease (Huddart et al., 2018). Inadequate primary health care in India and the prevalence of poorly qualified practitioners as primary points of care also hinders the timely referral to a TB detection center (Das et al, 2015). Among the patients who consulted a medical facility over the three months prior to the start of their TB treatment in Operation Asha’s centers, half went to see a private doctor, and 36 percent received an injection – a blatantly inadequate response to TB symptoms. Furthermore, almost half of the patients in our sample report fearing social discrimination if it became known in their community that they had TB. In qualitative interviews, many patients lacked knowledge about TB symptoms. They did not suspect that they had TB at their first consultation with a health provider and continued to cycle through different private and government doctors without suspecting TB. Early diagnosis and timely enrollment into the treatment, then, requires dedicated interventions on the part of public health systems, a need explicitly recognized as the first principle in the first pillar of WHO’s global “End TB Strategy” (2014).

The second challenge, treatment adherence, is important to prevent relapses and the development of drug-resistant strains of the disease, which are much more difficult and six times more expensive to treat (WHO, 2018). However, improving compliance is far from straightforward. To be cured, an infected patient needs to take a cocktail of up to seven drugs at a time, three times a week for a two-month period – and then once a week for a further four months. These drugs often have strong side-effects. Almost half of the patients in our study complained of severe physical side-effects like nausea, vomiting, fever, and stomach pains, while two thirds reported experiencing any side effects and a significant minority reported anxiety and dizziness. After about two months of treatment the symptoms of tuberculosis tend to wane off, but the side effects of the drugs remain. This leads patients to discontinue treatment early. Of the patients in our sample who were going through their second round of TB treatment, only about half had completed their earlier round. Interruptions occurred most frequently on the third month of treatment (median at day number 70), and most were due to a premature belief about recovery from the disease or to the significant side-effects of the treatment.
In response to the two challenges, increasing detection and ensuring adherence, India and other countries have adopted the Directly Observed Therapy, Short Course delivery method. In the DOTS model, local TB care centers are widely dispersed to facilitate patient access and the detection of new TB patients close to home. They hold medications on premises and require patients to come ingest each dose under direct observation of trained providers, who keep individual Treatment Cards recording daily pill intake and follow up on cases of noncompliance.

Another pillar of the DOTS strategy emphasizes the need to extend the reach of national TB programs through various possible delivery arrangements. In remote or overcrowded places where the national program itself is overburdened or inadequate, the strategy calls on governments to engage with the private sector in delivery. Private care providers include a vast array of actors ranging from unqualified practitioners to mission hospitals or NGOs, and the modalities of the partnerships vary according to the nature of the provider. Such arrangements appear to strengthen the state outreach capacity and improve TB-control outcomes, especially for vulnerable, hard-to-reach populations (WHO 2018).

While the implementation of DOTS has significantly improved efforts to contain TB over the last two decades (WHO, 2011; also see McMillen 2015 for a historical discussion), the system still suffers from several shortcomings, among which the level of monitoring and the quality of the record management system. The spread of DOTS centers does improve patient monitoring and decrease the effort and cost associated with visiting treatment centers, but not entirely: in our study sample, patients are on average three miles away from their treatment center, corresponding to 11 minutes of travel time. Visiting the center three times a week still represents a significant burden, especially for those physically debilitated by the disease. On the other hand, the high level of decentralization also reduces the TB Control Program’s administrative capacity to monitor health workers and ensure robust service delivery. Furthermore, the record system in those local facilities is entirely paper-based and treatment records are only collected by officials from the TB Control Program once every one or two months. NGO health workers can only monitor treatment adherence by cross-checking daily patient attendance against the full patient roster, which is time-consuming and leaves ample room for error and omissions. It also leaves room for opportunistic over- or under-reporting of events.

b. The intervention: TB primary care, health worker role and biometric monitoring

We conducted an experiment to determine if the installation of biometric monitoring devices in DOTS centers increased health workers’ effectiveness at ensuring treatment adherence of patients. We partnered
with Operation ASHA (henceforth OpASHA), a large NGO based in New Delhi, operating TB DOTS centers under public-private partnerships in India.

1. **TB care in the field**

TB is detected by examining sputum samples under a microscope to determine the presence of bacteria – a technique called *sputum smear microscopy*. In practice, symptomatic patients visit a Designated Microscopy Center, a medical unit generally located within a public hospital and leave a sample. Lab technicians perform the detection process. If patients are found to be TB-positive, administrative workers record them in a government register under a unique TB number and assign them to a DOTS center, depending on their location. The TB Control Program then sends a box of drugs bearing the patient’s name and containing the full treatment regimen to the DOTS center as well as a Treatment Card in which DOTS providers track treatment adherence by recording every visit made to the center. The patient can then start their treatment.

Patients take pills at their local DOTS center, which can be either a dedicated space or in a host facility such as a medical practice, a pharmacy, a store, a temple. Taking advantage of existing facilities saves operating costs to the NGO while ensuring that the DOTS centers stay open at least as much time as the facility is. Direct supervision of pill ingestion is then implemented by the DOTS provider, who can be either the health worker or another individual such as a pharmacist or shopkeeper, depending on the hosting facility. The DOTS provider records every visit on the patient’s personal Treatment Card, and declares Treatment Interruption, or default, if a patient misses pills for four consecutive weeks. Additional tests are done during the course of the treatment on sputum samples collected at the Designated Microscopy Center. At the end of the six-month period, DOTS providers record the treatment outcome on the Treatment Card and then report this information to government registers by handing the cards over to the local Designated Microscopy Center. Possible recorded outcomes include patient deceased, treatment completed, default, patient transferred out to another DOTS center, and patient cured.

2. **The role of health workers**

Health workers in our study are hired by our collaborating NGO, OpASHA, and attached to DOTS centers. In areas where OpASHA cannot establish a permanent treatment center, the organization hires mobile health workers who go around the communities and deliver information and treatment pills at home. Operations are supervised by one manager per city.

Health workers’ first objective is to promote early case finding. They canvass the area spreading awareness of TB symptoms and the availability of treatment, and asking ask health questions to households. When
they come across individuals showing TB symptoms, they provide information on the disease and the treatment process and encourages the individual to get tested for TB at the Designated Microscopy Center. When respondents are too weak to go to the Center in person, health workers collect sputum samples, which they deliver on their behalf. Some symptomatic individuals visit the Designated Microscopy Center of their own accord, often because they have prior knowledge of the medical care system or are referred by existing or former patients. In these cases, the interaction with the health worker starts only when they begin taking their treatment at the DOTS center.

Health workers are also tasked with tracking patients enrolled in the center who have missed a pill to bring them back onto the regular course of treatment. OpASHA instructs health workers are to review all Treatment Cards daily and identify if certain patients missed a pill. Missed pills should trigger follow-up by the health worker, either by phone or in person at the patient’s house. In qualitative interviews health workers describe strategies to boost medication adherence of defaulting patients, such as educating their families and neighbors to get their support in encouraging the patient, providing emotional support, increasing the frequency of house visits, or giving patients several doses at a time to spare them trips to the treatment center.

Health workers receive a base salary averaging Rs. 5600 per month (about 90 USD at the time of the experiment), with variations across cities. OpASHA measures the performance of health workers by seven indicators, including the number of new TB cases identified and the number of patients completing their treatment, as well as other indicators related to the maintenance of the treatment center. Qualitative interviews show that health workers perceived a strong pressure to achieve performance targets measured by the number of notifications and the number of defaults. Salary data show that health workers in all cities but three do receive financial incentives based on the number of new cases identified. Those disparities do not threaten the validity of results since treatment assignment was stratified at the city level. Three health workers – two in the control group and one in the treatment group – also received default-based incentives. Health workers commonly complain in qualitative interviews about the level of stress and their fear that poor performance would lead to salary cuts, sanctions, or contract termination.

3. **The biometric monitoring technology**

In partnership with Microsoft Research India, OpASHA developed eCompliance, an electronic biometric tool designed to help health workers monitor the treatment adherence of TB patients. The device consists of a tablet equipped with a SIM card and 2G internet data plan connected to a fingerprint scanner. A user-friendly application allows health workers and DOTS providers to navigate the system intuitively. The entire system costs less than USD 250 per center. In equipped treatment centers, all patients register their
fingerprints at the beginning of their treatment and scan them again every time they come to the center to take their pills. At the end of each day information on each new visit and pills taken is sent to the server through 2G or SMS, where a record of each patient’s profile and treatment history is kept. The health worker can use a second portable terminal to record home visits to patients. The system tracks all pill intakes in a simple adherence log. The interface allows health workers to view each patient's pill intake history, and to access the list of patients that have missed a daily dose and warrant some specific follow-up action.

The system also uses the health worker’s fingerprints for authentication and records the date and time when they log in and out of the system. In addition to a tool for health workers to monitor patients, it is therefore also a tool for NGO management to monitor effort and performance of health workers.

The data is stored in the terminal located in the health center and uploaded daily through mobile data to a server located in New Delhi for both backup and monitoring purposes.

### 3- Experimental set-up

#### a. Experiment and sample

The experiment consists of the random assignment of biometric devices to TB treatment centers. Given that patients must show up every other day to their treatment center, OpASHA sought to maximize the efficacy of its workforce by having one health worker operate 2 different centers open on alternate days, in accordance with the pill intake schedule of the patient population. We paired intervention areas in clusters of 2 covered by the same health worker, randomized at the cluster level, and stratified by city and type of health worker (fixed, mobile or hybrid health workers). Treatment areas were scattered and did not overlap. Two to four weeks after randomization, OpASHA installed biometric devices in centers or provided them to mobile health workers in the treatment group, and gave all health workers selected to use the technology a short training on the device.

The study sample includes 131 treatment areas grouped in 65 pairs spread across nine cities in four North Indian states – Bhopal, Gwalior, Indore, Sagar (in the state of Madhya Pradesh); East Delhi (Delhi); Korba, Raipur, Durg-Bhilai (Chhattisgarh); and Bhubaneswar (Odisha). The vast majority of these pairs (52) had one treatment center established in each of the two areas. Another 7 did not have any center and their 2 areas were operated by a mobile health worker, while 6 were characterized as hybrid, meaning that one of the two areas had a treatment center and the other did not. The experiment included a total of 85 OpASHA health workers. Figure 1 shows the geographical distribution of health workers in the sample: 50 are located
across four cities Madhya Pradesh, 21 in three cities in Chhattisgarh, 6 in Delhi, 10 in Odisha. The experiment lasted between 12 and 14 months, depending on the state.

Table 1 presents summary statistics for community health workers in the full sample used in the analysis (Columns 1 to 3). Health workers are 31 years old on average. Almost three quarters are men, 40 percent belong to the general caste and 81 percent are Hindus. They are well-educated on average, with a majority having completed secondary education or higher. While 75 percent of the health workers had previous work experience (for an average of 9 years) only 16 percent worked the social/NGO sector. A majority of health workers live in the neighborhood of one of their two DOTS centers, mostly in decent conditions: almost all of them have electricity and two thirds have access to tap water in their home. More than half own some land. They have some exposure to technology, 57 percent know how to use a computer and 40 percent have an email account. While some health workers expressed interest and motivation in social work in qualitative interviews, most mentioned doing this work because it is close to home and salaries are acceptable.

The full sample also includes 4911 patients recorded as undergoing treatment in one of the areas included in the experiment. Of those, a total of 3384 patients were surveyed at least once over the course of the experiment. Selected socio-demographic characteristics are reported in Table 2. Patients are men and women in almost similar proportions, 34 years old on average. They belong to the most deprived castes in India: Scheduled Castes and Scheduled Tribes (33 percent) and Other Backward Classes (36 percent). Only 18 percent belong to the better-off General category, as opposed to 41 percent of health workers. Most of them live in crowded, unsanitary conditions with high exposure to TB (see Section 2).

b. Data sources

We use many data sources in the analysis: several rounds of quantitative surveys administered to both patients and health workers, independent day-long monitoring of treatment centers, random spot-checks of treatment centers, qualitative interviews with patients and health workers, and program and administrative data from OpASHA and government registers. Apart from program and administrative data, all data were collected specifically for the purpose of this analysis.

1. Survey data

Patient surveys were administered to all patients enrolled in OpASHA treatment centers. Patients starting and finishing their treatment within the time span of the experiment received surveys at two points of their treatment cycle: an entry survey about a month after the beginning of the treatment (specifically between 2 and 6 weeks from the start date), and an exit survey within a month after the end of the treatment, whether the end was due to completion or interruption. Patients who were already undergoing treatment when the
experiment started were administered an augmented exit survey including socio-demographic information. Patients who were still undergoing treatment by the end of the experiment were administered an early exit survey.

The sampling frame was provided by the individual Treatment Cards maintained at each center. The survey team contacted all patients to set up appointments, unless the health worker warned them not to. Patients had the option to conduct the survey outside of their house if they wished to have privacy from their own family or community, or to refuse the survey. Most patients consented to being interviewed at home. Surveyors made three attempts to locate the patient before declaring them “Not found”. The completion rate is 67 percent; more than 35 percent of patients were surveyed twice, resulting in a database of 4600 observations from 3393 unique patients.

Health workers were scheduled to be surveyed three times: just before the installation of the biometric devices (baseline), then after about 6 months of intervention (midline), and finally at the end of the intervention 6 to 8 months after midline (endline). More than two thirds of them successfully completed all three, and 90 percent completed at least a baseline survey and one follow-up (either midline or endline). Health worker survey data include 219 surveys of 85 health workers. We systematically approached those who quit their job during the intervention and administered an additional module on the reasons for employment termination.

2. Observation days and random spot checks

Enumerators chosen from the survey staff conducted random daylong visits to treatment centers or to mobile health workers. They started the observation day by gathering the list of all patients scheduled to come on that day, then proceeded to collect information on attendance (both of the health worker and of each individual patient visiting the center), on activities carried out at the center and during home visits, and on the usage of biometric devices. Each center received an average of 2 random visits per month over the duration of the experiment, resulting in a dataset of 3096 observation days. We merged these data with the roster of patients in the experiment using three identifiers sequentially (TB number, lab number, and patient name) and with direct verification of each uncertain merge. The coverage of observation days is quite large. Overall, 85.6 percent of patients in the full sample appear in at least one observation day, and the average number of observation days is 3.7 per patient. This provides us with a thorough independent measure of medication adherence at the patient level.

To further strengthen the measurement of health worker attendance, senior survey staff conducted random, independent spot checks of the OpASHA centers, collecting additional information on whether the center was open and the health worker present. On average, each center received 7 spot checks over the duration
of the experiment for a total of 916 observations. These data provide a second independent measurement of provider attendance.

3. **Program and administrative data**

We also had access to center-wise data sheets, salary slips, Treatment Cards, and TB Registers. The center-wise data sheets are monthly reports prepared by health workers and used by OpASHA to track performance and outcomes and calculate salaries. They include data on the number of new patients enrolled, cases of noncompliance for each center or mobile area, and other monthly outcomes such as cases of patient transfers, deaths, and treatment failure. Salary slips include health worker-level performance data and the breakdown of monthly salaries.

Treatment Cards are standard documents provided by the TB Control Program and maintained by OpASHA health workers. They contain detailed information on the type of disease and the treatment start date, as well as daily treatment adherence and mid-course sputum test results. We were able to retrieve, digitize and successfully match individual Treatment Card information for 68.8 percent of all patients in the experiment.

TB Registers are the official source of patient-level information for the TB Control Program. They are kept by government TB officers at the central hospital and typically cover several Designated Microscopy Centers, listing the name and address of all enrolled patients, the dates and results of their initial and follow-up sputum tests, and the outcome of the treatment. The study team obtained access to TB Registers in all study sites except Indore (Madhya Pradesh). We retrieved, digitized and matched data using several identifiers in sequence and verifying unclear cases directly. Overall, TB register data were successfully matched for 94.1 percent of all patients in places where access was granted, allowing us to use the experiment to assess the extent of misreporting in official government registers.

4. **Qualitative interviews**

The study team conducted semi-structured interviews with 45 current and former OpASHA health workers and 47 patients. Open-ended questions to health workers related to their intrinsic motivation, effort reallocation, and data reporting. Questions to patients related to knowledge about TB, pathways to diagnosis, and experience with their current treatment.

Respondents were randomly sampled from the pool of all the health workers and patients who were part of the experiment at the time. To elicit candid responses, we complemented the sample with former health workers no longer employed by OpASHA.

Interviews were conducted primarily in Hindi by a pair of staff, a Research Associate and a senior field staff with relevant language competence and previous experiences in the project. Based on accepted
qualitative interviewing techniques and standard ethnographic practices, an interview template was designed for interviewing health workers around previously mentioned themes of interest, paying special attention to aspects such as question order (non-threatening to risky), nature of questions (“grand-tour” questions, structured questions, hypothetical interaction questions), and inclusion of numerous prompts to probe into open-ended questions. Interviews were transcribed, translated and coded using a qualitative data analysis software along major themes of interest, some pre-defined at the inception of the study and others emerging from the transcripts.

c. Internal validity

We present balance checks in Table 1 columns 4 to 8 for the subsample of health workers present at the onset of the experiment. Column 8 reports the p-value of the test that the difference between means in the control group and in the treatment group is null. Out of the 22 characteristics tested, five have different means across treatment and control groups (significant at the 5 percent level), which is slightly higher than what should be expected. Health workers in the treatment group are younger on average and more likely Hindu. They have more limited work experience and are less likely to have tap water in their housing. More importantly, they historically report a lower number of detections at baseline, i.e. of patients enrolled in OpASHA centers prior to randomization. In the analysis, we control for variables that show baseline imbalance.

Table 2 reports the balance checks for the sample of patients. To avoid potential selection biases in balancing tests, we restrict the sample to patients enrolled prior to the onset of the experiment (columns 4 to 8). Treatment and control samples display non-significant differences across observable characteristics, except for 2 out of 25, which would be expected statistically.

Among the 87 health workers who participated in the experiment, 66 were hired at the onset of the experiment and 21 joined in as replacements. Health workers left the organization for various reasons, including poor performance (48 percent), alternative work opportunity (24 percent) and illness (4 percent). Results show that the likelihood to attrite is not significantly affected by the treatment (Table A1 in Appendix, Panel A), and differences in observable characteristics of attritors (Panel B, columns 1-6) or replacements (columns 7-12) across groups are small in magnitude and statistically not significant.

The equality in retention rates across groups is consistent with the fact that biometric devices did not significantly affect health workers’ satisfaction with their jobs (we will return to this in Section 6 below). The logistics of replacing health workers also limited the scope for differential selection. The simplicity of the technology made it unnecessary for OpASHA to select particular profiles to operate centers equipped with the devices. One selection criterion for a replacement health worker included living in or near the
intervention areas. To allow for a rapid replacement of health workers terminating their contract, employee selection was based on a waitlist of existing ranked applicants. As a result, employee turnover was handled rapidly, with departing health workers being replaced on the day of their departure generally, and within 8 days on average. Anecdotal evidence suggests near to 100 percent acceptance rates by candidates who were offered the position.

4- Impact on TB control outcomes

We report impact estimates on two main areas of interest. First we first look at the impact of biometric devices on treatment adherence, which was the main TB control outcome targeted. We use various measures of adherence, rule out the existence of a confounding factor linked to (differential) selection of patients, and look into the impact on health outcomes. Results are estimated on outcomes measured independently by the research team. Second, we turn to the impact of the intervention on health worker effort, measured by attendance and intensity of activities. Several sources of independently collected data are used for this set of estimates.

All results are first presented in a non-controlled specification that only includes strata controls, and in a controlled specification that also includes the following set of covariates: gender, age, caste, religion, literacy, household size, house ownership, years spent in the area, and time to go to the center. We check the robustness of all results on patient adherence to restricting the sample to patients enrolled before the beginning of the experiment, since by construction their selection was not affected by the treatment, and display the results in Tables A3-A9 in Appendix.

a. Impacts on treatment adherence

Table 3 reports results on treatment interruption and adherence. Treatment outcomes are measured using independently collected surveys. The definition of default follows the WHO guidelines and reflects the situation in which a patient has stopped taking pills after 4 months of treatment or less or missed pills for a period of 2 months or more. Columns 5 and 6 show that the likelihood to default drops by 1.5 to 1.8 percentage points, depending on the inclusion of covariates in the regression, from a mean of 7.3 percent in the control group. This 25 percent drop is statistically significant at the 5 percent level, and its magnitude makes it highly significant from economic and public health standpoints too. There is no detectable impact on other treatment outcomes, including the death of patients over the course of the experiment, most likely due to the low fraction of such cases in the control group within the time frame of the experiment.
In Table 3 Panel B we report program impact on a series of individual measures of treatment adherence. Daily adherence to the treatment is measured routinely by the health worker through the individual Treatment Cards provided by the TB Control Program. In practice, the quality and accuracy of these cards vary tremendously and there is very limited capacity to ensure truthful reporting. The results presented here rather rely on the large number of independent observation days organized in all participating TB treatment centers. On such days, an independent enumerator would record all individual visits made to the center by patients or surrogates, and all visits made by the health worker to the patients’ homes. We then compute the adherence outcomes as the fraction of patients who did receive their pill on that day relative to the total number of patients who were supposed to. Columns 1 and 2 show that on an average day, the share of patients who came to the center in person increased by 12 percentage points from a mean of 54.2 percent in the control group, a 25 percent jump due to the presence of biometric monitoring technology in the center. The impact does not simply correspond to a substitution effect, in which patients who would normally send a trusted relative to the center to pick the pills now have to make it to the center themselves to scan their fingerprint. As columns 3 and 4 show, the impact on adherence is lower but still large and significantly positive after including the possibility that patients send a relative. Columns 5 to 8 include pills given during home visits, completed or announced by the health worker. Note that those outcomes better capture the variety of pill delivery methods, but they may well be inflated by health workers who may overreport their home visits – especially in the control group where monitoring is less tight. Results show that the significance of the impact holds after including all pill delivery methods. If there was any substitution between delivery methods, with an increase of patient visits and possible decrease in other methods, it did not offset the overall impact on treatment adherence. As shown in column 8, on an average day 81.1 percent of patients in the control group would receive their pill by any method available, including potential inflation by the health worker. This number goes up to 89.3 percent in the treatment group, a 10 percent hike.

Table 3 Panel C leverages patient surveys to confirm impacts on compliance with some key features of the DOTS protocol defined by the WHO. The share of patients who declare sending some other person to pick their pills for them, a breach of the DOTS protocol, is massively reduced by the introduction of biometric devices. From a mean of 36.7 percent in the control group, this share drops to 14.9 percent in the treatment group, a 21.8 percentage point drop. As discussed above, this reduction does not offset the positive impact of the intervention on patient adherence. Similarly, there is a large reduction in the share of patients who declare picking pills for one week or more (i.e. three or more separate doses of medication) and would then not be observed ingesting them. These results show that the biometric monitoring device is an effective instrument for improving adherence to the DOTS protocol.
b. Impacts on detections and patient selection

The sizeable impacts noted on treatment adherence could potentially harm the other key objective of TB control programs, early case finding or patient detection if, for instance, health workers reallocate effort toward default-prevention activities to the detriment of detection activities (for detrimental multitasking effects, see Holmström and Milgrom, 1991). There could also be strategic decisions by health workers to limit patient notification and keep their caseload reasonable for medication adherence activities which they know will be monitored more stringently.

Table 4 reports the impact on detections – Panel A, on the average number of verified patients detected per center per month. As explained in Section 3, verified patients are ones surveyed by independent enumerators in face-to-face interviews. All patients reported by health workers in experiment areas were part of the survey sample. Although up to three attempts were made to conduct the survey, a fraction of actual patients could not be interviewed for reasons such as the lack of accuracy of addresses registered in the program data, the difficulty to identify particular dwellings in urban slums or the fact that work or other obligations may make patients unavailable despite efforts to connect at a convenient time. As we will show in Section 5, a fraction of these patients were also not verified because they never were true TB patients.

Panel A shows that there is no impact on the number of verified patients detected by health workers, which can be interpreted as a rejection of the multitasking hypothesis. It also suggests that there is no strategic quantitative adjustment of the number of detections by the health workers to facilitate future default-prevention work. We also examine whether health workers focused their detection efforts on patients who may be less prone to defaulting in the future. For this we focus on the subset of patients enrolled after the beginning of the experiment. Differences in observable characteristics would suggest differential selection of certain profiles by health workers when biometric monitoring is in place. Panel B of Table 4 displays balance tests for this subset and shows no observable difference between treatment and control.

c. Impact on patient’s health

We use a comprehensive health section from the patient surveys to test whether the improved adherence to treatment induce measurable improvements in patient’s health. Table 5 Panel A reports results on physical health, measured by a general perception score, reported difficulty in physical activity, or an index of health symptoms. Panel B reports results on mental and emotional health, including optimism, future-orientation, distress, and happiness. Overall, we do not find any significant impact of the intervention on patient health.

These results are perhaps not surprising, given that the biometric monitoring technology focuses on preventing defaults from treatment, a dramatic but relatively rare occurrence. Patients whose health was
most likely to improve because of the intervention are the ones who would have interrupted their treatment otherwise. They constitute a small fraction of the overall sample, with default rates around 7 percent in the control group.

**d. Impact on provider effort**

We now examine whether the positive results identified on TB control outcomes can be accounted for by impacts on health worker effort. We measure effort using two independent data collection methods: first, the observation days during which an independent enumerator would record centers’ opening and closing time and health worker presence (in addition to keeping a log of visiting patients as analyzed in Section 4.b); and second, the random, unannounced spot checks.

Table 6 Panel A shows results on health worker attendance, measured using both sources of data. The intervention did not have a detectable impact on whether or not the TB treatment centers were open (columns 1–4 and 11–12), reflecting the fact that they are often situated in another facility operating at regular hours irrespective of DOTS activities. However, the biometric device did increase the presence of the health worker at the center. The controlled regressions consistently show a statistically significant 5 to 7-percentage-point increase in the likelihood that the health worker is present at the center (columns 5 and 6 for observation-day and 13 and 14 for spot-check data). With presence during an average observation day at a fairly high 83.4 percent in the control group, the marginal improvement is sizeable. Absence was cut by about 27.4 percent according to observation-day data and 17.6 percent to spot-check data. Furthermore, observation-day data allow us to compute rigorous attendance durations by subtracting breaks taken by the health worker during the day. The data point to significant improvements in presence time at the center, by about 20 minutes per day or 18.8 percent (columns 7 and 8). These results echo qualitative findings on health-worker perception of their work with the biometric devices. Most explain (and complain) that the installation of the devices forced them to stay longer at the center to make sure that all patients could come and scan their fingerprints, lest a missed-dose alert would be recorded in the system. As columns 9–10 show, based on observation-day data, biometric tracking is complemented by more frequent visits by OpASHA staff members, which suggests greater enforcement of work requirements by the management structure.

We now look into the impact of biometric devices on the intensity of activities performed by the health worker. Table 6 Panel B reports results based on a variety of measurement sources. Using observation-day data, columns 1–2 show that biometric monitoring induces a higher frequency of home visits to patients to ensure that pills are not missed, which health workers intensify by about 50 percent (6.1 percentage points from an average of 11.6 percent in the control group). This increase comes in addition to higher attendance
figures at the center, pointing to a sizeable improvement in overall effort. It cannot be attributed to a variation in the proportion of mobile health workers across groups since the randomization of biometric equipment was stratified on this variable.

The other dimensions of intensity of activities do not seem to be affected by the presence of biometric monitoring devices. Health workers do not report intensifying activities aimed at preventing defaults (columns 3 and 4) or reducing detection activities (columns 5 and 6), consistent with the absence of impact on the number or characteristics of new patients detected. The patient surveys confirm that health workers do not seem to intensify or reduce the service they provide to them in response to the biometric devices (columns 7–10).

5- Impact on data quality

a. Qualitative evidence on misreporting

Salary bonuses linked to performance indicators or general pressure from management may lead health workers to forge performance numbers. Our structured qualitative interviews documented that misreporting can take many forms. Most of the qualitative evidence was offered by former or retired health workers, who were more open about this part of their experience than current employees who generally offered more perfunctory responses.

On detection figures, health workers report instances where patients were entirely fabricated. While patients are supposed to go to the Designated Microscopy Center themselves to provide sputum and get tested, health workers can offer to carry the sputum to the Center. In some instances, they used sputum from one infected patient to create one or more additional cases, which were then registered under fake names or using names of noninfected individuals in the community for greater realism. The ghost patients would then be assigned a TB number, and a box of pills would be delivered to the DOTS center. Health workers could also create fake patients without tinkering with sputum samples by colluding with the lab technician at the Designated Microscopy Center. Qualitative work revealed instances when several ground-level agents agreed on a mechanism for artificially inflating detection numbers and splitting the associated monetary incentives. There was reluctance by certain health workers to grant access to patients for survey purposes, a reaction that more forthcoming health workers (especially retired ones) interpret as reflecting the fear that some of their patients would be identified as fake.

While health workers had evidence to share on fake detection numbers, misreporting default data is not as clearly documented in the qualitative interviews. Respondents tended to identify this issue as relevant at
the hospital level rather than at the health-worker level. In the regular reporting system without the biometric devices, there is, however, scope for health workers to misreport defaults. It only takes a check in the patient’s Treatment Card to report continued adherence to the treatment. Patients who interrupt their treatment can also be marked as having moved or having passed.

Biometric devices may dramatically reduce the scope for misreporting data. By requiring fingerprint scans, they cut the possibility that entirely fake patients are enrolled and shown as adhering to the treatment. The biometric system does not fully eradicate any possibility of forgery. Qualitative evidence revealed cases in which health workers managed to detect fake patients through one of the methods described above, and then have a relative or neighbor register their fingerprint at enrollment and scan them over the course of the treatment. It is however a much costlier strategy.

b. Measurement

Health workers report the number of detections and defaults for each treatment center to OpASHA every month. For month \( i \) and treatment center \( j \), the reported outcome \( \tilde{Y}_{i,j} \) can be broken down as follows:

\[
\tilde{Y}_{i,j} = Y_{i,j} + M_{i,j}
\]

where \( Y_{i,j} \) is the true outcome and \( M_{i,j} \) corresponds to misreporting. Let \( \bar{Y}_T \) be the average monthly outcome reported for centers in the treatment group and let \( \bar{Y}_C \) be the average monthly outcome reported for centers in the control group. The treatment impact estimated on reported data can be decomposed as follows:

\[
\bar{Y}_T - \bar{Y}_C = (Y_T - Y_C) + (M_T - M_C)
\]

In equation (2), \((Y_T - Y_C)\) is the “true” impact, which we computed based on an independent data and reported in the previous section. \((M_T - M_C)\) is the difference in misreporting between the group of centers equipped with biometric devices and the control centers. Given that health worker and treatment center characteristics are balanced across intervention groups, any unintentional dimension of misreporting is most likely evenly distributed and would be canceled out in the difference. It is worth noting that the outcomes of interest are little susceptible to random mistakes, as notifying a new patient or concealing a defaulting patient involves several steps over a period of time (see Section 3b). Since we cannot exclude that health workers still find ways to misreport in centers equipped with biometric devices, the quantity \((M_T - M_C)\) is a lower bound for the actual level of misreporting happening in centers that are not equipped. Yet it might be close to the reality, as \(M_T\) is expected to be small: as discussed above, the scope for misreporting is much reduced in areas equipped with a biometric device, where patients need to scan their fingerprint at the time of their detection and several times a week over the entire course of their treatment.
Equation (2) shows that the impact coefficient estimated on program-generated data can be interpreted as the sum of the “true” impact and the difference in misreporting between centers equipped with biometric devices or not. Conversely, to measure the impact of the intervention on misreporting, we measure the difference between “true” impact estimates based on outcome measures from our independent data (surveys, observation days and spot checks) and estimates based on program and government data (salary sheets and productivity reports, Treatment Cards, and TB registers).

c. Impact on misreporting

Table 7 reports impact estimates on default as computed using different data sources. Columns 1 and 2 show the impact on the likelihood to default measured by independent survey data and correspond to the coefficients previously displayed in Table 3 (Panel A, columns 5 and 6). Columns 3 and 4 show the impact on defaults as measured in the Treatment Cards maintained by health workers in each treatment center, which were collected and digitized for this study. This primary source of program data shows no difference in the likelihood to default between the center equipped with the biometric device and those that are not. This indicates that health workers underreported defaults in centers without a biometric device and that the extent of underreporting decreased in centers equipped with a biometric device. Health workers without a biometric device were able to artificially lower the number of defaults in their reports, with the effect of concealing the actual decrease in default rates in the treatment group. This finding obtained from Treatment Card information holds in other official program data. The center-wise data sheets produced by OpASHA (columns 5 and 6) appear to reflect the underreporting of defaults by health workers without a biometric device. At the end of the reporting chain, government TB registers display the same bias (columns 7 and 8). To test the significance of the difference in coefficients we use a seemingly unrelated regressions framework to compare treatment effects on actual likelihood to default and on default as reported in the Treatment Cards, in the program data, and in the government registers. Results displayed at the bottom of Table 7 confirm that we can reject coefficient equality.

We now turn to measuring effects on misreporting of detections. Table 8 shows the impact of the monitoring tool on the number of new cases reported. Panel A columns 1 and 2 report the reliable estimates already shown in Table 4, which are based on program data cross-checked with patient surveys, and show that the biometric monitoring did not have a significant impact on the number of detections. Columns 3 and 4 use OpASHA data compiled from the center-wise data sheets. The numbers reported are aggregate numbers of new patients and the unit of observation is a health center × month. Note that in the control group the survey teams were able to verify about 42 percent of all reported notifications (1.58 out of 3.77 in an average month). This percentage reflects the difficulty of survey operations in urban slums (which apply to the intervention areas too), but also the extent of overreporting detection figures, a factor that would be mostly
cancelled out by the biometric monitoring. Results displayed in columns 3 and 4 show that the number of reported detections is significantly lower in biometric-equipped centers than in regular centers. With about 0.8 fewer notifications officially reported in equipped centers, and a null impact of the intervention on the number of “true” patients newly detected, the monitoring technology appears to have curbed overreporting of detections by about one fifth of all patients declared.

Another way to identify fake patients is to look at the likelihood that the survey team could verify each individual patient across the treatment and control areas, a strategy similar to the matching-based test used to measure changes in the number of “ghost” program beneficiaries (Niehaus and Sukhtankar 2013, Muralidharan et al. 2014, Banerjee et al. 2016). Panel B reports survey outcomes for the full sampling frame of patients attempted to be surveyed and shows that patients were significantly more likely to be found and surveyed in areas equipped with a biometric device, and significantly less likely to refuse to answer (columns 1 and 2). These results are consistent with the forgery mechanism described in the qualitative evidence in which health workers enroll noninfected individuals as patients. Knowing that enumerators were visiting every one of their patients, health workers would have talked the fake ones into refusing the survey. It also appears quite consistent with the fact that certain health workers and supervisors resisted survey operations or tried to undermine them over the course of the experiment. Another possible overreporting mechanism consists in making up individuals and registering them under fake identities and addresses. Those ghost patients would not be found by surveyors. However, ghost patients would only add to the pool of patients that went unlocated due of the difficulties of finding individuals in poorly mapped slums. This might explain why the coefficient attached to the survey outcome “not found” has the expected negative sign but turns out to be non-significant.

Finally, Panel C of Table 8 confirms that the misreporting of patients carries over into official government registers. As OpASHA works by delegation of each Indian state to implement the national TB control program, its data management system forms an integral part of the public data. We were significantly more likely to find patients from the treatment areas in the official TB registers, suggesting that biometric devices filter out fake patients that would otherwise be present in public health records.

Taken together these results provide robust evidence on the existence of forgery in health records at the level of the primary provider and also in government registers. The installation of biometric monitoring devices appears to significantly curb the level of misreporting and improve the timeliness and truthfulness of health records. Accurate and timely data are a critical component of the state capacity to design and implement policy, identify needs and allocate resources accordingly, or respond to crises. This paper provides even clearer evidence on the importance of data quality for policymaking: the results from this impact evaluation would have been quite misleading if the existing official data had been used to measure
outcomes. As shown in Tables 7 and 8, the study would have reached the conclusion that biometric devices had no impact on the likelihood to default while reducing the number of patients notified by health workers – and may have been used to recommend abandoning the technology. The ability to take evidence-based policy decisions is a component of state capacity that relies heavily on accurate program data.

6- Acceptability, sustainability and scalability

Increased monitoring may be met with strong resistance by primary providers, to the point of not being sustainable (Banerjee et al. 2008). This section presents evidence on the acceptability of the equipment by health workers and patients, and its sustainability in treatment centers.

Table 9 presents the impact of the intervention on key measures of health worker wellbeing. The intervention resulted in fewer (over-)reported detections (Table 8), even though the actual number of detections remained stable (Table 4), which in turn led to reduced monetary compensation associated with detections. Salary data obtained from OpASHA shows the reduction in salary. Columns 1 and 2 confirm that the compensation for detection activities was significantly reduced in, resulting in a 6 to 7 percent cut in the total monthly salary for health workers using the device compared to the control group. In addition, qualitative evidence shows that health workers bemoan the loss of discretionary power, for example that they can no longer decide to provide several doses of medicine at the same time or give the drugs to a family member for convenience.

However, health workers in intervention areas enjoy a significantly reduced workload. In the control group 22 percent of health workers report an excessive workload. The installation of biometric devices reduces this share by half, or even virtually solve the issue depending on the specification (columns 7 and 8). This reflects the fact that the devices make it easier for health workers to do their work on default prevention. The coefficient attached to the number of challenges faced in ensuring treatment adherence is negative, although non-significant. Qualitative evidence shows that health workers are satisfied with having less reporting responsibility since all their activity record is accessible by management directly from the server. They also appreciate having all patient details immediately available, including their adherence record, treatment phase and corresponding dosage. They acknowledge that automatic alerts make the tracking of patients easier, but also regret that scanning fingerprints adds to the number of steps involved in distributing drugs and forces them to stay longer hours at the center to make sure all scheduled patients have had a chance to visit.
Overall, despite imposing additional constraints on both health workers and patients, the installation of biometric devices does not seem to impact the satisfaction level of either. On the health worker side, we do not see any significant impact on their job satisfaction – if anything the coefficients have a positive sign (columns 9 and 10). This result might reflect the fact that although the job offers less scope for fraudulently seeking compensation, it is overall easier and more rewarding – an idea consistent with the very similar levels of turnover observed in the treatment and control group (Table A1 in Appendix). On the patient side, satisfaction levels are remarkably similar across treatment and control centers. Patients across groups are equally satisfied with healthcare received and equally likely to recommend OpASHA to other symptomatic individuals. This may result from two opposite impacts: patients who are receiving more frequent home visits may be more satisfied, while those under increased pressure to come to the center in person may be less satisfied.

The acceptability of biometric devices may account for the durability of the equipment in the centers. The information collected on observation days includes whether the equipment (fingerprint reader, laptop, USB key) is present in the center and working properly. Taking advantage of the duration of the experiment we can then estimate the evolution of technology presence over time. Table 10 shows that in the first month of the experiment 92 percent of intervention centers did have a biometric equipment present on the day of observation. Columns 1 and 2 report how each passing week affects this likelihood and show that the presence of technology is remarkably stable over time. The remaining columns suggest that there may be a slight reduction in the presence of certain peripheral components over time, but the graphic representation of the evolution (Figure XXX in Appendix) shows a stabilization at high levels after 6 months.

Not only is the technology present and functioning over time, the impact on the main outcomes of interest appear to be maintained. If the impact of the technology was driven by a novelty factor, or if health workers had found ways to circumvent the technology over time, then the positive impacts on TB control would only manifest themselves in the early weeks or months. Table XXX in Appendix presents coefficients associated with an interacted term treatment × time, in addition to the direct impact of the treatment itself. The time variable is obtained by computing the number of days lapsed between the installation of the biometric device and each patient’s treatment start date (for outcomes measured through patient surveys) or between the installation and the biometric device and the observation day (for adherence outcomes measured through observation days). Coefficients associated to the interacted term are consistently not distinguishable from zero, which provides reassurance that the impact is sustained over time.

These findings suggest potential significant benefits of scaling up this technology in the Revised National TB Control Program (RNTCP). In 2012 the Government of India launched a web-based centralized database of all TB patients called Nikshay, based on the digitization of paper-based Treatment Cards. By
aggregating TB records from across India in one unique dataset, *Nikshay* serves as a dashboard for the National TB Control Program. Yet as it relies on the digitization of paper-based documents, *Nikshay* is not set up for identifying duplicates or establishing unique identity. The quality of the database is entirely reliant on the quality of information on the paper-based Treatment Cards, and the digitization process may even add a layer of potential error in the recording system. Further, the digitization process is implemented with a lag of about one to two months between when the patient misses a dose and when the corresponding information gets entered in the electronic database, thus disabling any quick corrective action to ensure patient compliance.

On the other hand, biometric identification technology is increasingly available in India, especially since the national roll-out of *Aadhaar*, the Government of India’s ambitious program to provide a biometrics-based Unique Identification (UID) to all residents of India. *Aadhaar* is a 12-digit individual identification number which can be authenticated through fingerprint scan, iris scan, and registered cellphone number and serves as a proof of identity and address, anywhere in India. Currently *Aadhaar* covers about 1.2 billion individuals, corresponding to about 90 percent penetration among the adult population. The objective of the Government of India is to expand the system enough for various social programs to use it as a platform for service delivery. The eCompliance technology we study appears to perform well as a point-of-care device and a robust interface for leveraging biometric identification in TB care provision. It is worth pointing out the anecdotal evidence that portable devices were successfully used by mobile health workers, even though their number was too small for us to conduct separate impact analysis. It appears that the technology is not restricted to brick-and-mortar facilities but is conducive to use in remote settings, rural areas or other hard-to-reach places.

With *Nikshay* and *Aadhaar*, the infrastructure is in place for using a biometric recognition technology linked in real time to an integrated database of care beneficiaries, which would massively improve the reliability of records, enable rapid follow-up on patients who miss pills, and facilitate the rigorous implementation of financial incentive schemes or Direct Benefits Transfers to both providers and patients (RNTCP 2017). The Government would have information to make rapid-cycle adjustments to program implementation and hold implementers accountable for results, which would also facilitate the efficient engagement with private providers for greater progress toward eradicating the disease.

7- Conclusion

This paper provides original evidence on the benefits of using biometric technology to strengthen frontline service delivery and improve the reliability of government data. We exploit the random placement of the
technology across four Indian states and use a large variety of data sources to obtain a detailed understanding of impacts and mechanisms and uncover the extent of misreporting in the sector.

Results point to a sharp reduction in treatment interruptions, greater medication adherence, and a greater compliance to the requirements of the DOTS protocol defined by the WHO. These results are accounted for by a higher level of attendance by health workers and greater effort dedicated to following up with defaulting patients through more frequent home visits. Potential downsides did not materialize: we do not see a reallocation of health worker efforts away from other critical activities such as early case finding or a strategic refocusing of care toward more compliant patients. In addition to these impacts on direct TB control outcomes, the biometric technology dramatically reduces the scope for forging health registers. Triangulating different data sources, we are able to estimate the extent of underreporting of treatment interruptions and overreporting of new cases detected in the government registers. The fact that the technology is a productivity-enhancing tool helps explain why it appears well accepted by both patients and health workers – despite imposing more efforts on both – and why its impacts are sustained over an extended period time. As a result, strengthened identification and embedded monitoring appear conducive for scale-up through India’s TB Control Program.

The technology studied here measurably improved state capacity at two levels. First, it increased the capacity of public programs to efficiently extend their reach by addressing the principal-agent problem inherent to the decentralization of service provision. It enhanced the quality of implementation of a delivery model based on partnerships with private sector actors in charge of establishing highly localized points of care. This resulted in a large impact on the likelihood that isolated, vulnerable people receive adequate treatment against a deadly disease. Second, the technology improved the accuracy of public-health data, which increases the capacity of the state to identify needs and allocate resources accordingly. Accurate data pave the way to adequate decisions: in our case, official data affected by misreporting issues would have masked the positive effects of the program and likely led to its abandonment.

In order to enhance state capacity in a durable way, changes need to be accepted and sustainable. The technology we study improves the productivity of health workers while at the same time holding them more accountable for their performance, and thus may hold lessons for strategies to improve state capacity beyond the important case of TB treatment. Sustainable improvement in the commitment of government agents may be facilitated by a simultaneous effort to build the state’s legitimacy and improve the work conditions of its agents. Financial incentives can only go part of the way toward meeting this challenge. Financial incentives can only go part of the way toward meeting this challenge. Technology may improve state capacity, not only as a tool to deliver services smarter and faster, but also though a positive impact on
government workers, if they willing to accept greater scrutiny in exchange for a simplification of their daily tasks.
8- References


Mohanan, Manoj, Katherine Hay, and Nachiket Mor. 2016. "Quality Of Health Care In India: Challenges, Priorities, And The Road Ahead." *Health Affairs* 35 (10): 1753-1758.


