The Punjab Model of Proactive Governance: Empowering Citizens Through Information Communication Technology[1]

Findings from an Early Review of Evidence

Michael Callen
Department of Economics
University of California San Diego

&

Ali Hasanain
Deparment of Economics
Lahore University of Management Sciences

July 25, 2011

Disclaimer: This review is intended to support an ongoing effort by the Government of Punjab to calibrate the design of the Punjab Model of Proactive Governance to achieve maximum effectiveness. It is necessarily based on limited data and at this stage the program has not been evaluated using a methodology that would permit us to make conclusive statements about program effects. The discussion here is intended to provide a review of the program to date and to provide independent guidance going forward. We make no claims regarding the causal effect of the program on corruption at this stage.

[1] This project is a collaborative exercise involving many people. Sheharyar Ahmed, Asim Fayaz, Abdul Rehman Naeem, and Zubair Bhatti provided valuable assistance with understanding the program and the data. We are thankful to seminar participants at the World Bank for useful feedback. Eli Berman, Zubair Bhatti, Ali Cheema, Naved Hamid, Asim Khwaja, Craig McIntosh, Ijaz Nabi, Naseer Rana, and Jacob Shapiro provided useful discussions. All mistakes are our own. The opinions expressed are those of the authors and should not be attributed to the World Bank, the Lahore University of Management Sciences (LUMS), or the University of California San Diego (UCSD). For further information, please see the Punjab Model website: http://www.punjabmodel.gov.pk/

1. Executive Summary

The Punjab Model represents the novel application of Information Communications Technology (ICT) to engage citizens and to close the space for extortion in the delivery of public services. The program has three objectives. First, the program seeks to deter corruption by monitoring petty officials through large-scale solicitation of service beneficiary feedback. Second, it seeks to promote direct citizen engagement and thereby signal commitment to service provision on the part of the government. Last, it seeks to improve services by allowing citizens to report problems.

What is the Punjab Model?

The Punjab Model deters corruption by collecting data on bribe-taking by bureaucrats who administer basic services (e.g. property registration, the licensing of drivers, providing glucose drips). This happens in three basic stages. In the first stage, government offices providing basic services record beneficiaries' mobile phone numbers along with critical transaction details and transmit them using either the internet or a novel SMS-to web technology to a central database. Next, numbers are called and enquiries about corruption made. The current scale of the Punjab Model requires a call center make calls in order to contact a reasonable percentage of beneficiaries, but this is unnecessary for smaller operations. A possible variant at this stage is to send mass SMS messages to all phone numbers and to ask them to text back feedback directly or to screen individuals for a follow up phone call. In the third and final stage, the data are aggregated and analyzed for patterns consistent with corruption. If many cases are reported against a particular official or office, the relevant official may be given a warning, be suspended, or if need be, dismissed from service.

Innovations and Cost-Effectiveness

Theoretically, the program should impact service delivery by increasing the probability that attempts at extortion and under-performance in service delivery are detected. This blunts incentives for corruption and also provides citizens more bargaining power when dealing with officials who might ask for a bribe. It has the highest chances of success for transactions in which the corruption is not collusive (i.e. those where the citizen in more interested in reporting corruption than in sustaining the relationship or in continuing to receive the

benefits provided in return for a bribe). Additionally, the model should target bribes taken by low-level bureaucrats, so that the benefits of taking punitive action for the more senior official outweigh the potential costs of removing someone with political power or connections.

The logic of the program is supported by considerable theory and evidence from the economics of corruption. ICT provides a powerful means of removing information bottlenecks that allow officials to underperform and to request bribes. The program is suited to viral adoption, requires a limited upfront capital outlay, and requires only cellular connectivity for all technological aspects of the program to function. These features of the program combined with the rapid proliferation of mobile phones and especially of SMS messaging in Pakistan suggest great promise for the program to scale quickly and at comparatively low cost. Notably, the reliance only on existing cellular networks: (i) reduces the cost of capital and maintenance relative to smartphone and internet-based ICT interventions; (ii) reduces reliance on internet penetration and large-scale electrification; and (iii) makes the program user-friendly for both beneficiaries and monitored bureaucrats.

A programmatic innovation of the PM is to channel information to mid-level government officials who have the right career advancement incentives to reduce corruption, but lack actionable information. This increases the potential for the program to result in positive remedial action and to be sustained—which has proven to be a major challenge for comparable interventions in South Asia—despite objections from potentially corrupt bureaucrats to increased monitoring.

Some Quantitative Facts Relevant to the Program and Early Progress

- The program is currently operating in 15 districts in the Punjab and the Chief Minister has officially endorsed a scale-up and constituted a Cabinet committee to oversee full implementation of the program.[2] As of June 05, 2011, 30,941 transactions in health; 463 driver's licenses; 583 character certificates; 1,582 education pension disbursements, and 51,258 property registrations were recorded by the program.
- We encode 370 citizens' feedback SMS responses for the evaluation and find that 82 out of 370 respondents are explicitly thankful that the government has undertaken this initiative and that 161 reported a positive experience at the office being investigated.

- Use of SMS messaging has also increased dramatically in recent years in Pakistan. From August to September 2010, just under 49 Billion text messages were sent (The Nation, 1 April 2011). SMS tariffs are also incredibly low. As an example, ufone offers a package which allows unlimited SMS to be sent for Rs 666 annually or about 8.30 USD.[3] Rapid growth rates in SMS use, and the prospects for a 3G network, make this trend likely to continue.
- Data from health call feedbacks provide clear evidence of the program's ability to document deficient service provision in a large sample at low cost. Of 428 connected calls in the health sector, about 38 percent report individuals are only getting some medicine and 28 percent are getting no medicine.

Refinements and the Way Forward

In this assessment we argue that future progress depends on answers to three questions. Can the program provide citizens the correct incentives to report bribe-taking truthfully and accurately? Can the program be set up in such a way as to ensure that the numbers reported by officials for return calls and SMS contact are accurate? Will the information result in punitive action if patterns of corruption are documented?

Providing citizens with proper incentives to truthfully report dissatisfaction with service provision is critical to the program fulfilling its objective of identifying patterns of underperformance. Our report recommends a mixture of socializing the program, guaranteeing the anonymity of individuals who report corruption, and providing evidence that reporting corruption results in socially beneficial action. We also recommend that the ability for officials to influence which numbers are received by the call center be diminished.

We believe the nascent program is intuitively appealing and potentially highly scalable and cost-effective. So far the focus has been on developing the technology and accumulating the political support necessary for the program to operate. We recommend in the next stage of development taking full advantage of the data generation capabilities of the program to carefully calibrate the program using a smart design approach. This maximizes the chances for success of the program when it operates at provincial-level scale across many areas. Additionally, this program is well-suited to rigorous randomized control

[3] http://www.ufone.com/prepaid_sms.aspxRapid Prices accessed July 7, 2011

evaluation, which our team aims to undertake with the program principals in the coming months.

2. Introduction

The global expansion of access to ICT and especially of access to cell phones is enabling a range of new development innovations. Mobile banking (Jack and Suri, 2010) and mobile applications for monitoring elections (Callen and Long, 2011) provide clear examples of how ICT can improve development outcomes by mobilizing information.

The Punjab Model represents an especially innovative and promising application of ICT to a critical social problem—petty corruption. In the Punjab, according the Multiple Indicator Cluster Survey (MICS), some 70 percent of households have a cell phone with 29 percent of the poorest quintile and 60 percent of the second poorest quintile having phones. Use of SMS messaging has also increased dramatically in the past three years. From August to September 2010, just under 49 Billion text messages were sent in Pakistan (The Nation, 1 April 2011). For example, ufone offers a package which allows unlimited SMS to be sent for Rs 666 annually or about 8.30 USD.[4] Rapid growth rates in SMS use, and the prospects for a 3G network, make this trend likely to continue.

Theories of corruption (Becker and Stigler, 1974; Rose-Ackermann, 1975), which have since received robust empirical support (Di Tella and Schargrodsky, 2003; Olken, 2007) provide the common sense argument that increasing the probability that corruption is detected reduces its incidence. More recent arguments also highlight the importance of separating corrupt officials from the institutions that are supposed to police corruption (Callen and Long, 2011).

By putting the citizen s in charge of policing and reporting corruption, the Punjab Model dramatically increases the number of individuals that must be coordinated, relative to traditional anti-corruption task forces, for corrupt individuals to co-opt and undermine the anti-corruption effort. An additional argument, which supports the potential effectiveness of the program, is that it breaks the monopoly of local officials on local information. The career advancement benefit to cracking down on petty corruption for mid-level and senior officials, in the current context, should outweigh the costs of dealing with upset subordinates and possibly of losing a small cut of the rents. Moreover, this initiative gives senior bureaucrats and politicians easy access to a very large amount of real-time data on service delivery. The low cost of

[4] http://www.ufone.com/prepaid_sms.aspxRapid Prices accessed July 7, 2011

collecting this data, coupled with a new-found ability to make comparisons across all districts may change the political incentives these senior officials face: it may be profitable to focus on providing better quality service, which can easily and reliably be verified, rather than cultivating patronage.

As one District Coordination Officer (DCO) points out, low ranking officials can only extort citizens if they can do so without being noticed by their superiors or if they will not be punished because they are in a collusive relationship.[5] This provides a clear logic for using ICT—it reduces the costs of quickly aggregating the information necessary to detect misbehavior—and prescribes a specific target—officials who are not in collusive relationships with their superiors.

The Punjab model has three central objectives: (i) reducing petty extortion; (ii) empowering citizens to hold government officials accountable through proactive engagement; (iii) improving service delivery by facilitating feedback on quality. In the next section, we review arguments that pursuing all three objectives should have important positive follow on effects for economic growth, human development, and citizens' support for the government. Increasing citizens' support for the government may represent an especially important impact, as this is increasingly viewed as a critical input to growth and human development (World Development Report, 2011).

This paper is structured as follows. Section 3 links the program to current research on anti-corruption and governance strengthening. Section 4 provides a brief history of the program and describes how the program operates practically and conceptually. Section 5 puts the Punjab Model in historical and international context. Section 6 draws on previous studies, theory, and data generated by the program to try to predict impacts and to suggest refinements. Section 7 provides qualitative evidence of the program so far and Section 8 outlines the future of the program, makes recommendations, and concludes.

3. Background: Corruption, Citizen Welfare, and Political Legitimacy

3.1 Corruption, Citizen Welfare, and Political Legitimacy

[5] Rural Health Center (RHC) operations, Medico-Legal Certificates, property registration, domicile issuance, and the provision of driving licenses provide examples of services that are either free or at nominal official cost or with clearly defined taxation, but that are commonly subject to extortion because they are difficult for supervisory officers to observe due to their volume and geographic dispersion.

Bribe payments have been shown to have a convincing negative effect on firm growth (Fisman and Svensson, 2001), school enrollment and academic achievement (Reinikka and Svennson, 2005), and also to increase the capture of public resources (Khwaja and Mian, 2005). There is also some mixed cross-country evidence the corruption is related to reduced macroeconomic growth (Mauro, 1995; Knack and Keefer, 1995; Svennson, 2005). More immediately relevant for the Punjab Model, Bertrand, Djankov, Hanna, and Mullainathan (2007) study the provision of driver's licenses in Delhi, India and find that corruption creates consequential misallocations in government-provided services. In this study, 71% of driving license recipients did not take the licensing exam and 62% were deemed unfit to drive by an independently administered test calibrated to be the same as the government test because individuals could pay a bribe to avoid taking the test. The total fee for obtaining a license, in this case, was Rs 1,120 or 2.5 times the official fee, and results in clear undermining of regulations to guarantee safety by licensing only competent drivers. Similar to property registration in Pakistan, facilitating agents who fall outside the jurisdiction of government regulations on official corruption, were responsible for collecting and passing on bribes. A related cost, identified by Acemoglu and Verdier (2003) is that corruption harms economic performance by drawing talented people from productive occupations into the economically unproductive corruption sector. To the extent that the Punjab Model is successful in removing the rents to ineffectual public administration, these problems will be diminished.

Increasing the accountability of public servants is also likely to benefit citizens as this will increase their ability to influence the services they receive. This has been clearly documented in the case of increasing accountability for elected officials. Fujiwara (2010) finds that the introduction of an electronic voting technology which facilitated voting for less educated Brazilians increased enfranchisement and shifted spending toward a pro-poor policy (public health) resulting in more pre-natal visits by less educated mothers and ultimately in a reduced occurrence of low-weight births. Also in Brazil, Ferraz and Finan (2010) show that mayors with re-election incentives misappropriate 27 percent fewer resources than mayors without re-election incentives in the presence of random auditing. Besley and Burgess (2002) show that state governments provide more relief expenditure and public food in response to floods and major reductions in food production in Indian states where electoral accountability is greater. Last, Chattopadhyay and Duflo, (2004) provide evidence that political reservations for women caused more investment in clean drinking water and in roads in Rajasthan and West Bengal. The clear lesson that emerges from these studies is that empowering citizens and increasing the accountability of public officials can considerably improve governance and,

consequently, citizen welfare.

The final potential benefit, which has received less attention but is perhaps the most relevant, is that governments which are accountable are more likely to be popularly supported, and so may be more stable and less likely to suffer political opposition. Broadly, the Punjab model promotes legitimacy first by increasing the accountability of potentially corrupt officials, second by improving service delivery, and finally by having high ranking officials call citizens, thereby advertising the government's concern.

Eliminating corruption is only part of developing institutional legitimacy, but legitimacy as an objective is entirely consistent with the implementation of the Punjab Model. Mobilizing citizens and removing the feeling of fatalistic helplessness in the face of corrupt petty bureaucrats is a core objective of the program. Along these lines, in the introduction to the most recent World Development Report, World Bank President Robert Zoellick argues that "Institutional legitimacy is the key to stability." This echoes earlier arguments made regarding the source of continued opposition to the government in Afghanistan: "Widespread corruption and abuse of power exacerbate the popular crisis of confidence in the government and reinforce a culture of impunity" (McChrystal, 2009). The Punjab Model seeks to challenge the culture of impunity in some very corrupt offices. The World Development Report makes two prescriptions, which are strongly supportive of implementing innovative anti-corruption and citizen-empowerment programs. It recommends that:

> [To] break cycles of insecurity and reduce the risk of their recurrence, national reformers and their international partners need to build the legitimate institutions that can provide a sustained level of citizen security, justice, and jobs.

> [To increase legitimacy and inclusion governments must] provide information to citizens and mechanisms for legal recourse to resolve disputes and complaints, including complaints against the state.

Collectively, this evidence suggestions that the Punjab Model holds great promise to improve welfare and increase the popularity of the government by providing an effective means of holding officials accountable. The low-cost and easy portability to other sectors, provinces, and countries, also provide good reason to believe this is a highly cost-effective and scalable intervention. The literature, however, provides some guidance on potential vulnerabilities of the program, which can possibly be remedied through careful calibration and smart design.

First, almost all corrupt behavior requires contracting within networks of senior level and lower level bureaucrats. For example, McMillan and Zoido

(2004) using detailed and comprehensive data on bribes paid by Vladimiro Montesinos Torres, the secret-police chief for Peruvian President Alberto Fujimori, show that bribes paid by Montesinos were larger for bureaucrats and institutions who had information or authority that might imperil Fujimori's corrupt network. While the Punjab Model targets corruption at a much lower level, the lesson from this example is that an anti-corruption program targeting low level officials might just reduce their bargaining power relative to more senior officials in the chain. Officials subject to Punjab Model monitoring may just need to hand over a larger share of their bribes as protection against information from the cell calls being used as a basis for prosecution. The extent to which this is true depends on the whether the rents from involvement, through a network, in petty corruption are more valuable than the potential for career advancement or political gains that comes from clamping down on corruption.

Second, as we discuss more thoroughly below, even petty officials will try to undermine the Punjab Model if the costs or likelihood of getting caught for doing so are low. Callen and Long (2011) find strong evidence that corrupt election officials working on behalf of candidates to rig the 2010 election in Afghanistan responded quickly and using several different means to try to recover votes that were lost because of the introduction of a new monitoring technology. We similarly find speculative evidence of officials subject to monitoring by the Punjab Model shifting approaches from providing spurious numbers to providing duplicate numbers in response to a program change, which we detail in Section 5.

The Punjab Model holds great promise for improving citizen welfare and popular support for the government by providing citizens a powerful means to demand accountability from their public officials. Corrupt officials and their associated networks, however, are resilient, adaptive, and have strong incentives to preserve their rents. Below, we provide recommendations on approaches to implementation that will help to overcome what we believe to be the most important challenges for the future of the program: (i) circumventing attempts to undermine the program, particularly officials selectively providing numbers for follow up and (ii) providing citizens with correct incentives to honestly report on bribes and on the quality of services received. Some of the problems of citizens underreporting corruption may be alleviated as citizens are made more aware of the purpose and intention of the program. To date, the program has operated without any mass advertisement.

3.2 Civil Services in Pakistan

Despite a series of reforms, civil services in Pakistan are still largely patterned on the system inherited from the British Raj. Under the system, the administrative affairs in each of the 36 districts in Punjab province are run, largely autonomously, by a DCO. The autonomy is a remnant of the Raj; a highly federal approach was necessary because transaction costs prohibited a high degree of central coordination.

Historically, DCOs are usually given a large residence, from which they conduct affairs. The residence and the office are quite insular, despite efforts to open up. The DCO of a typical district may be supervising more than 10,000 government employees and citizens may number some 2 million. Citizens often have to bear considerable financial and time costs in order to gain an audience with the DCO, and the disconnect between the DCO and his constituents is vast.

The widespread public anger against perceived endemic corruption, a suddenly and increasingly vibrant media, what appears to be an increasingly competitive democracy, and the penetration of the cell phone industry in the country provided an environment naturally conducive to the use of ICT for improving governance. Set in this institutional and historical context, the component of the Punjab Model which provides DCOs with a set of citizens numbers to call today represents a remarkable departure from current and traditional governance.

The reason that this works is because the program provides senior officials[6] and elected officials with a tool to fight corruption that has a self-contained mechanism to compel active championing—it provides a direct connection to citizens and so a direct means to signal a proactive means to improve governance. While a lot of the credit for pushing the program through the early phases belongs to a team of reform-minded officials in the Chief Minister's Secretariat and in the districtswe believe that the program has achieved the scale necessary for it to be naturally supported and championed by more politicians. The fact that the Model increases outreach and provides citizens a feeling of being heard makes it a natural fit with the incentives of a politician as elections loom closer.

4. The Punjab Model: What It Is, How It Works, and How it Was Developed

[6] Senior field officers -- Commissioners, who may supervise three to six DCOs, and the DCOs - are drawn from federal and provincial cadres with promising career trajectories and are considered less likely to be entrenched in local politics.

The Punjab model requires government officials who provide services to record the cellular number of beneficiaries and a few pieces of critical information as part of the receipting process. These data are then passed on to local officers and to a call-center through an online data entry form or through SMS messages. Beneficiaries are then contacted via SMS or by a direct phone call and asked about the transaction. The purpose is to proactively engage the citizen, rather then waiting for complaints, and to remove the information bottlenecks that corrupt officials can exploit to extort citizens. As such, the Punjab Model is not designed to impugn officials based on single bits of evidence. Rather, it is designed to diagnose patterns and use them as a basis for remedial action. A remarkable feature of the program is the speed at which it generates data on the quality of service provision to pursue these objectives. To our knowledge, no comparable data collection initiative exists in any developing country.

4.1 How It Works: The Punjab Model as a Management Innovation

The purpose of this subsection is to provide detail on the operation of the program. We provide this background first because it is critical to understanding how the program works to reduce corruption and second because the innovation is easily applied in other settings. We hope this section provides guidance on application for interested readers.

The core advancement made by the program is to systematically collect the cell phone numbers of service beneficiaries in order to have an independent agent, usually a supervisory officer, contact beneficiaries and investigate whether illegal bribes are being taken. Two points bear emphasis. First, the infrastructural requirements to implement the core idea are minimal. Indeed, the earliest incarnation of the Punjab Model, the Jhang Model, required only paper, pencil, a cellular phone, and a manager interested in data on bribe-taking by his subordinates. It was first implemented by a mid-level official, in one of Punjab's 36 districts without any supporting ICT infrastructure; he simply required his subordinate officers involved in property registration to write down the cell phone numbers of beneficiaries and then provide the numbers to him so that he could call a random subset and investigate bribe-taking. This approach is broadly applicable and only requires that supervisory officials ask their subordinates who transact with citizens to record cell phone numbers for parties to the transaction. Second, having the supervisory officer pro-actively call service beneficiaries removes a cost that is monetarily small but might be psychologically large, increasing the volume of feedback relative to a complaint line. Second, this approach is likely to provide a more representative sample of service users and is also less likely to receive spurious claims.

To understand how the model operates, it is useful to go through the process step-by-step:

- In the first stage, a citizen goes to a government office to obtain a service (for example, to register a newly purchased property). During the transaction, the office will record the citizen's mobile phone number along with critical transaction details. The details will vary from sector to sector, but broadly the data collection should assist with verifying the identity of the citizen (e.g. CNIC number, name), the time and place of the transaction, and details that might be useful for identifying corruption during the transaction (e.g. official duty, price paid for the property). An important refinement to the procedure that has been made is the introduction of a receipt once the information is provided. This seems to improve compliance because the citizen's lack of knowledge about the project becomes less important.. The data collected are then transmitted, often using sms to web technology, to a data collection center. The fact that the Punjab Model does not exclusively rely on data entry at computer terminals has many advantages: the cost of capital and maintenance is lower; power and network outages do not cause delays; and the difficulty of learning to key in an entry for an official using their existing phone is often far less than the difficulty in undertaking computer training.
- In the second stage, the mobile numbers are then passed on to supervisory officers so that the numbers can be called and enquiries about corruption made. The current scale of the Punjab Model requires a call center make calls in order to contact a reasonable percentage of beneficiaries, but this is unnecessary for smaller operations. In the current model, a random subset of numbers are also forwarded to DCOs and other senior officials. Having senior authorities call may have a stronger deterrent effect. Box 1 provides some guidance based on experience from the Punjab Model on how to write scripts to encourage individuals to provide honest information about their experiences during the transaction. A possible variant at this stage is to send mass SMS messages to all phone numbers and to ask them to text back feedback directly or to screen individuals for a follow up phone call, asking them to SMS "1" if they were asked for corruption and "2" otherwise. Robo-calls should also be considered as a potentially cost-effective option both for contacting citizens and for gathering data on transactions. The program designers view these various forms of communication simultaneously as survey devices and as tools for politicians and bureaucrats to communicate with the citizenry.

- If a large volume of data is collected during stage 2, then it needs to be systematically recorded in a database, and patterns need to be analyzed. In addition to direct reports of bribe payments, other proxies should be considered. Duplicate number entries, invalid number entries, numbers where the respondent does not recall engaging in a transaction, or cases where individuals report a total transaction costs (including payments to agents such as waseeqa nawees) greatly in excess of the duties that were reported to be received.
- If patterns which clearly suggest corruption attributable to one individual or to one office appear, remedial action should be taken. The primary aim of the exercise is to create deterrence through communication and citizen empowerment. Research shows that transparency of information flow itself reduces corruption. Punitive actions are however necessary to create a credible threat. If many cases are reported against a particular official or office, the relevant official may be counseled, warned, suspended, or if need be, dismissed from service.

Some Practical Suggestions:

The experience of the Punjab Model to date indicates some practical design issues, which influence program effectiveness. First, the model is much more likely to have an effect if transactions do not involve collusion between the bribe-taker and bribe-payer in misappropriating a government service. For example, there is a high demand for spurious Medico-Legal Certificates (MLCs) so that fraudulent police cases can be filed. The state official is clearly misappropriating a government service to benefit the bribe-payer, and so the payer is unlikely to report corruption as he does not want to reveal his complicity. Second, the punitive capacity of the bribe-taker needs to be carefully considered. If the official and the citizen are in a repeated relationship, then the official can seek retribution by excluding the citizen from future services. Alternatively, even in one-off transactions, if the official has powers that can create serious trouble for the citizen, such as the case with police officers, then the model is likely to be ineffective.

The use of phone calls versus text messages to gather feedback involves some trade-offs. For SMS to be effective, literacy rates need to be high, or responses need to be extremely simple (e.g. respond "1" for corruption, "2" otherwise). Phone calls are much more expensive, but provide a few benefits in addition to being suited to populations with low literacy rates. They also allow more senior government officials to directly interact with citizens. Last, if the phone calls are structured to be conversational, it is possible that individuals

might open up and feel confident responding honestly. Robo-calls represent an additional option.

A related issue arises when beneficiary population uses multiple languages. Response rates for the Punjab Model increased when calls in Sariaki were made in areas where Seraiki is the dominant language. For highly heterogeneous populations, it is best to provide phone calls and texts in all relevant languages.

The program currently works in the education, police, health, and revenue sectors. As of June 05, 2011 numbers corresponding to 30,941 transactions in health; 463 driver's licenses; 583 character certificates; 1,582 education pension disbursements, and 51,258 property registrations were

BOX 1: Designing Scripts to Encourage Accurate Reporting

The evidence in this section was developed during experimentation at by staff at the Pakistan Telecommunication Company Limited (PTCL) call center, which has been contracted to implement the most recent phase of the Punjab Model during early April 2011. The lessons that emerge from this are:

- Starting scripts with an indication that the call is coming from the government makes respondents nervous and is confusing. It is better to start by verifying the identity of the citizen.
- Indicating that the call is associated with a particular government office can help reassure the respondent. It is best to associate the call with an office that has authority but is not local and so is less likely to be interested in the affairs of the citizen for some other reason.
- An open-ended question such as "were there any issues when you availed yourself of the service," will allow respondents to vent frustration. Key information should be picked out from this response. This may also put the citizen at ease before sensitive questions about fraud are asked.
- Be clear about the purpose, but be concise. Too many references to authorities, punitive consequences, and corruption can be distracting and unnerving for the citizen.
- Asking about corruption in the third person, such as "did you see any individuals paying bribes" did not appear to work.
- It is useful to back out corruption by asking about specific itemized fees and then asking about the total expenditure. Probing about the difference can provide information about bribes paid to agents or other intermediaries.
- The length of time that elapses between the transaction and the call involves tradeoffs. Calling immediately means that individuals are more likely to recall details of the transaction. However, if the transaction will involve future follow up transactions, as is the case in property registration, then individuals may not provide information about corruption. A delay may allow them to complete follow up transactions and then they can comfortably report corruption without fear of retribution during later transactions.

recorded. The program currently operates in 15 of the 36 districts constituting the province, namely: Bahawalnagar, Bahawalpur, Faisalabad, Gujranwala, Gujrat, Hafizabad, Jhang, Khanewal, Mandi Bahauddin, Nankana Sahib, Narowal, Rahimyar Khan, Sahiwal, Sheikhupura and Sialkot.

In implementing the program, telecommunications regulations, especially regarding mass SMS messaging need to be considered. In this case, the project designers have included a declaration of consent to be contacted on the form where respondents are asked to record their mobile numbers. In this way, the program is able to accord with the Pakistan Telecommunications Authority's regulations on spam SMS messaging.

An additional novel feature of the program is the method of using SMS-to-Web technology in order to provide an on-line real time record of service provision, which represents a fantastic management innovation to DCOs and other senior officials. Importantly, it records and instantly centralizes key data on all transactions and facility specific feedback in the sectors and districts in which it operates. This is currently operational in the health sector, providing a "citizen's report card" of a Rural Health Centre, that can be used by more senior officials to take stock of performance. This component has incredible potential to replace expensive and labor-intensive service recording systems in settings where money, human capital, and other capacity for such work is thin.[7] The infrastructural requirements for this system are much less than for traditional paper and pen approaches to keeping accounts on services delivered or receiving feedback.

Bringing the Punjab Model to Scale

The Punjab Model, at its core, is a very simple innovation that can be implemented at the level of a single district, but, as we review in the history section, in the last year the program has scaled very rapidly. The 15 districts it currently covers represents a large portion of the Punjab with a considerable population. As with scaling any intervention, this necessarily creates an additional host of managerial and logistical challenges. For example, the quality of interaction, when a DCO is making a few calls and has an earnest interest in assisting his beneficiaries, is likely to be much better than that from a call placed by a disinterested calling center agent. Systematizing data collection,

[7] A recent innovation is to provide individuals with a copy of the receipt which collects the core details to be transmitted to the supervisory official or to the call center. This permits additional verification that the contact information and transaction details collected by the transacting bureaucrat are transmitted accurately up the chain.

management, and analysis is also more complicated when the program operates at scale.

These challenges can be overcome through a process of experimenting with different methods for managing large call volumes and large amounts of data. As the program is brought to scale, provisions should be made to be sure to sufficiently resource the program to allow careful training and management of call center staff and to permit best practice data management and analysis. While the speed of scaling and the large number of beneficiaries now served is remarkable and the achievements should not be understated, making sure that sufficient time and resources are provided to permit the program to be well-implemented when it operates on a province-wide scale is critical. Given the exceptional promise of the program, we also recommend that the program be provided resources and official sanction to begin experimenting in new sectors as well.

History

The program has its beginnings in Jhang district. Zubair Bhatti launched the first effort to use cell phones to systematically record attempts at extortion in property registration when he was DCO in charge of the district. . Specifically, he requested the Deputy District Officers in charge of signing off on property transactions to record the mobile numbers of citizens who completed the transaction. This is reviewed in the Economist story reprinted in Box 2.

On July 05, 2008, the News International, an English language publication with nationwide circulation, documented the story. A few days later, CM Sharif recommended that it be scaled province-wide.

In early 2010, the program migrated to the Chief Minister's Secretariat, and began to receive active technical support from the Punjab Information Technology Board.

On April 30, 2010, the Chief Minister's Secretariat issued an official communication indicating that a committee, to be overseen by the Secretary of Implementation and Coordination, would be charged with overseeing a district-wide roll-out. The program was designated to scale in Revenue (Registration of property documents), School Education (pension and leave cases of education officials), Health (MLC, small surgicial procedures, and the provision of free medicines to emergency patients in government hospitals). The project was additionally to be piloted in the Excise and Taxation department. There was additionally a requirement in the early phases that the program be piloted in Excise and Taxation. By June 2010, the system was scaled across 10 districts (The News, June 16 2010).

In 2011, with the active personal interest of the Chief Minister, a special cell was set up in the Chief Minister's Secretariat to pursue the program and in June of 2011 a committee was established to oversee the scale-up of the program to a provincial scale. In January 2011, the program was one of 13 Innovation Fund Challenge recipients out of a field of 170 applicants awarded by the World Bank Group. The prestigious award signals the esteem of a group aware of the leading development innovations across the world. The World Bank wrote that *it looks forward to "partnering with the Punjab Government — and learning from its experience — to develop, scale and institutionalize this proactive outreach to citizens for improved governance*."

Box 2: Economist Article Eureka Moments September 24, 2008

Mobile phones can also be used to root out corruption in more direct ways. For example, Zubair Bhatti, a Pakistani bureaucrat, asked all clerks in the Jhang district who handled land transfers to submit a daily list of transactions, giving the amount paid and the mobile-phone numbers of the buyer and the seller. He explained that he would be calling buyers and sellers at random to find out whether they had been asked to pay any extra bribes or commissions. When charges were subsequently brought against a clerk who had asked for a bribe, the others realized that Mr Bhatti meant business, and buyers and sellers reported a sudden improvement in service. Mr Bhatti extended the scheme to other areas, such as cracking down on vets who demanded bribes from farmers, and has proposed that the Jhang model, as it is now known, be adopted in other districts. "It could easily be institutionalized with a call centre," he says. "It could have big vote-getting influence."

This program is novel in that it is the first example of a government-sponsored effort to harness the power of ICT to remove information bottlenecks that challenge the successful supervision of government officials. To achieve the scale necessary to have an impact, and for the monitoring to result in action that meaningfully improves governance and service provision, the championship needs to be sustained at the the highest levels. Willingness to use a novel technique is not without risk; truly reform-minded support is required for success.

Conceptual Operation of the Program

Theoretically, the program should impact service delivery by increasing the probability that extortion and under-performance is detected. This blunts incentives to be involved in or condone corruption for government officials and also provides citizens more bargaining power when dealing with officials who might ask for a bribe. It has the highest chances of success for transactions in which the corruption is not collusive (i.e. those where the citizen in more interested in reporting corruption than in sustaining the relationship or in continuing to receive the benefits provided in return for a bribe). For the model to work, it must target bribes taken by low-level bureaucrats, so that the benefits of taking punitive action for the more senior official outweigh the potential costs of removing someone with political power or connections.

The proactive engagement of citizens reveals to them that the government is serious about addressing their concerns and making sure that they receive satisfactory services. This is likely a reliable vote-getter and should create a short run increase in support for the government. Positive support and genuine legitimacy, however, are necessarily long-run concepts. They derive from citizens believing that, over the long run, the services and benefits of social order provided by the state outweigh the taxes, divestiture of personal autonomy, corruption, and other costs associated with the prevailing political system (Lake, 2006). In this view, for the Punjab Model to be a true success and to have a sustained effect on citizens' perceptions, it must follow through on its promises of increased engagement, more official accountability, reduced corruption, and higher quality service provision. The way to maximize the likelihood of this outcome is through careful honing to guarantee effectiveness.

As we discuss in Section 6, data generated by the Punjab Model provides the opportunity to identify patterns consistent with corruption. This point bears emphasis. The program is not designed to be a complaint receiving mechanism. Instead, it actively seeks information on service delivery by calling a random sample of beneficiaries. In this way, it creates a more scientific assessment of the quality of service delivery and removes some of the transactions costs, which have historically been a major weakness of anti-corruption programs.

As mentioned, there are two general challenges for the data to achieve this objective. First, data on cell numbers for each transaction must be accurately recorded. Officials who benefit from bribes and the ability to shirk should be expected to take advantage of the ability to distort this information, especially if they can do so without fear of consequence. The Punjab Model starts in an unusually advantageous position in working to redress this issue.

Detecting this behavior is not a problem; we document it below. Designing a measure to deter officials from misreporting is the next step. Second, citizen beneficiaries must have the right incentives to tell the truth when contacted. This may require familiarity with the program, belief that their identifying information will be held in strict confidence, an absence of concern that they might suffer retribution[8], and also some personal gain to reporting (even if the benefit is just the "warm glow" of knowing that services may improve for others). Knowing which of these conditions are necessary to achieve honest reporting is the natural next step, in our view, to designing the program to achieve further success. The program, because it generates data at an exceptional rate, is well positioned to experiment with different approaches towards solving the problem of officials fudging numbers and beneficiaries being unwilling to report corruption.

It is likely that for citizens to honestly report paying bribes, the motivation of the program, and the guarantee of anonymity for respondents, must be public knowledge. To date, publicity for the program has been limited to the few articles and the editorials cited in this report. The June endorsement of the Chief Minister's cabinet includes an initiative to greatly expand public awareness.

5. The Punjab Model in Historic and Comparative Context

Anti-corruption programs date back at least to the 13th century. The Ducal palace in Venice, in use during the 13th century, has a stone with a hole in it, through which people could inform the Duke about corrupt tax agents (Tanzi, 2000). Serious efforts to use telephonic communication to reduce corruption in developing countries began on a large scale in the 1990s. In Uganda, for example, a telephone hot-line was set up to enable people to report corrupt tax officials. For this a reward was provided of 10% of the recovered tax (*The Economist*, July 17th 1996:38). In 1995, the Mexican government, under President Ernesto Zedillo, undertook a major five-year governance strengthening reform called the Program for the Modernization of Public Administration (Programa para la Modernización de la Administración Pública or PROMAP) as part of the National Development (Plan Nacional de

[8] The more complaints are received about a certain office, the lower the ability of an officer to pinpoint complainants and exact retribution. However, whistle-blowers are usually those who speak up when others are silent, and so we consider anonymity, and the belief in insulation from retribution that it brings, to be very important.

Desarrollo or PND). This program involved a very large package of reforms, including an anti-corruption hot-line.

The founding program documents speak to its proactive anti-corruption stance:

> It is important that citizens have access to several mechanisms to express their opinion about government performance, such as opinion polls, suggestion boxes, surveys, interviews, opinion groups or direct consultation with representative citizen organizations…as users become familiar with the criteria used to carry out official acts and the mechanisms for providing services to the public, subjectivity will be avoided *and corruption will be eliminated.* …Whenever possible, agencies and offices should directly and systematically consult the target population they serve or, failing this, do so through social organizations, chambers or representative organizations, in order to know their needs and specific proposals, by means of: … electronic mechanisms to record and analyze complaints and accusations….
>
> (http://zedillo.presidencia.gob.mx/welcome/PAGES/library/od_publicadmon.html)

In the end, Zedillo's anti-corruption program and the associated reform effort largely resulted in failure. It is argued to have done so because full transparency imperiled rents available to senior officials (Arellano Gault and Guerrero Amparán, 1998). As mentioned, this is distinct from the Punjab Model. First, taking remedial action against the officials monitored by this program is not a problem for supervisory officers. Second, the initiative, on some level, is citizen-based. Consequently, it has a degree of insulation against interference by high-level political interests.

In recent years there has been a proliferation of ICT-based efforts to improve governance, reduce program linkage, and improve service delivery. In India, ICT is being used to monitor the presence of health service providers in rural health facilities. It is also being used to allow rural health professionals to consult with experts in cities to get a second opinion on diagnoses (Muralidharan, In Progress). Also in India, USAID Development Innovation Ventures (DIV) has sponsored a program to coordinate reporting on the spread of new diseases to improve epidemiological forecasting. ICT has played a role in crowd-sourcing, especially using Ushahidi, to coordinate disaster response and to monitor elections in Mexico, India, Sudan, and Mexico. In Afghanistan, ICT is currently being piloted to reduce leakage in infrastructure delivered under the $1 bn flagship community-drive development (CDD) program, the National Solidarity Program (NSP), which now operates in over 22,000 villages across Afghanistan.

6. Predictions of Program Effect from Previous Studies, Theory, and Data

Evidence from Previous Studies

Previous studies provide at least four lessons regarding the optimal design and potential effectiveness of the program. First, attention to the program needs to be sustained, and the information it generates needs to be the basis for periodic action. Second, officials who are taking bribes should not be able to influence the data being brought in by the program. Third, the program needs to evolve and to continually renew and recalibrate itself based on the data it generates. Last, the program needs to have a careful understanding of the set of considerations that influence whether an official adheres to the program or attempts to undermine it.

Di Tella and Schargrodsky (2003) study a crackdown on corruption in the public hospitals of the city of Buenos Aires during 1996-1997. They find that the crack down had a well-defined, negative effect on prices that hospitals reported paying for basic, homogenous health materials. Prices fell by 15 percent during the first 9 months, but increased subsequently to only 10 percent less than the pre-crackdown price. In this study, they find that higher wages reduce corruption when the probability of that corruption is detected is at intermediate levels, but has no effect when the crackdown is at its maximum intensity. The lessons for the Punjab Model from this policy experiment are twofold. First, unless the increase in the probability that corruption is sustained, it is reasonable to expect reversion. This means that implementers of the Punjab Model need to remain vigilant and be sure that the information bottlenecks it eliminates are not replaced by others. Below, we provide evidence that, even though the technological innovations provided by the Punjab Model make it possible to observe every transaction, there are still other actions bureaucrats can take to disguise misbehavior. Second, if the Punjab Model only raises the probability of detection for misbehavior from zero to intermediate values, then other incentives, such as wages, will still hold sway, and are likely to be *more* relevant to the corruption calculus than when detection probabilities are zero.

Perhaps the best-known study of anti-corruption effort is the work by Olken (2007) documenting that public audits reduce theft in road construction programs in Indonesia. He also documents the poor performance of community-based initiatives to reduce corruption. A fundamental problem faced by these efforts was that the roads constructed in Indonesia were contracted directly with communities, making it difficult to separate beneficiaries from the individuals who benefit from graft. A point in the study

that receives less attention than the core finding, but which is potentially instructive for the Punjab Model, is that the only cases in which community monitoring was effective were cases in which anonymous feedback on corruption was solicited via village schools, completely bypassing the village government and so preventing village elites from soliciting feedback only from their supporters. The lesson for the Punjab Model is that monitoring will be most effective when the monitor and the corrupt official are separate entities, so that it is possible for reporting corruption to be compatible with the monitor's incentives. Specifically, it is likely that part of the problem with invalid numbers, duplicate numbers, problematic contact rates, and an implausibly low number of responses reporting corruption, has to do with the monitored being responsible for recording numbers. Delinking these two functions may have positive effects on both respondent truth-telling and increasing the accuracy of recorded numbers.

The Punjab Model, to our knowledge, is completely novel in using cellular technology and web-technology to police corruption and is the only state-sponsored effort of this sort. There are, however, two comparable efforts to police election fraud using mobile devices. The first, documented in Aker, Collier, and Vicente (2009), provided individuals during the 2009 Presidential Election in Malawi with cell phones and asked them to text in problems with malfeasance. These individuals were later interviewed, and it was found that those who were instructed to look for corruption felt that there had been more problems. Their impact on corruption, however, is not clear. Another example is from Callen and Long (2011) who deployed monitors to 465 or 7.6% of polling centers during Afghanistan's 2010 Wolesi Jirga election. The technology worked by announcing to polling center officials that any discrepancies between the votes cast and tallied at a given polling center and those which appeared in the center-person specific accounts that comprise the national aggregate would be detected. The most instructive finding in this research for the Punjab Model is that there is clear evidence that candidates responded rapidly to this technology by attempting to rig using approaches that could not be detected by the technology and which they were not aware were being measured. This suggests a careful analysis of the data generated by the Punjab Model to observe patterns of substitution and also of potentially gathering additional data on bribes collected in ways that could not be detected by this model. For example, if officials can dissuade individuals from reporting corruption, by, leveraging some potential threat of retribution, then calling individuals will be insufficient to observe corruption. There are two possible solutions. Institutionally, making sure that officials do not retain identifying information for beneficiaries, such as phone numbers, makes retribution impossible. As a data issue, independent collection, which is increasingly done

by hiring surveyors to avail themselves of a service, will allow bribes and retribution threats to be observed.

Ferraz and Finan (2008) study the effects of a random audit of federal funds transferred to municipalities in Brazil. They find that mayors found to be corrupt were electorally penalized by the public most in municipalities where local radio was present to divulge information. The lesson here is that the public will react to reliable information about corrupt behaviors, and that the method of dissemination is important. The results of audits were pro-actively broadcast in Brazil, resulting in a large penalty for corrupt mayors. The effect of the Punjab Model on official behavior, this evidence instructs, will depend in large part on how the data it generates are used.

One of the most powerful features of the Punjab Model is its capacity to generate real-time data on the quality of service provision especially through the Dashboard which aggregates sms and calling agent feedback. For example, in Gujranwala, 39.38% of contacted RHC users report receiving no medicine. This is a powerful statistic, which is informative for the EDO-H and for other stakeholders. We now turn to a preliminary analysis of this data to draw further lessons for program design. If there is any concern that official follow up will be insufficient, it may eventually, after a sufficient period of piloting, be beneficial to make the data generated by the Punjab Model public-facing.

Evidence from Data

Table 1 summarizes the key indicators that can be useful for supervisory officers, managers, and other stakeholders.

Table 1: Punjab Model Indicators

Pre-Call		Call	
Indicator	Use	Indicator	Use
1. Invalid Number	Data Entry Error; Malfeasance	1. Wrong Number	Data Entry Error; Malfeasance
2. Duplicate Number	Data Entry Error; Malfeasance	2. Disposition	Data Entry Error; Malfeasance
3. Service Price	Malfeasance	3. Fictitious Entry	Data Entry Error; Malfeasance
		4. Unsuccessful Response	Data Entry Error; Malfeasance
		5. Unsuccessful Calls	Data Entry Error; Malfeasance
		6. Corruption	Malfeasance

	Reported	
	7. Non Cooperative Staff	Quality of Service
	8. Received No Medicine	Malfeasance; Quality of Service

Data from the Punjab Model can be classified as either pre-call data and call data. Pre-call data are information that come in about the service, either via web entry or via SMS, which include several critical bits of data including the beneficiaries' cell phone numbers. We provide some preliminary analysis of this data below. Call data are data returned during the call or relayed from beneficiaries via SMS. These data, which can be visualized and investigated in real time via the SMS-to-web dashboard, are especially instructive about problems in service delivery.

Figure 1 plots the percent of total weekly property transactions that are either duplicate cell numbers or are numbers that are invalid. Duplicate numbers are those that appear more than once in the property registration, where numbers belonging to the same individual are dropped to remove legitimate multiple purchases and number. Invalid numbers are numbers that either: (i) begin with a prefix not assigned to cellular numbers in Pakistan; (ii) are all zeros; (iii) or all 1s. Around March 18, 2011, the system was transitioned from Punjab Information and Analysis Unit (PIAU) to Pakistan Telecommunication Company Limited (PTCL) and then to Dashboard entry, after which time clearly invalid numbers were no longer entered. The top line depicts the percentage of weekly numbers entered into the database that are duplicates. Above the weekly average points, the total number of transactions for the week is reported. Note that invalid numbers, in almost all cases, are a strict subset of duplicate numbers, before they are eliminated from the database.

Four observations emerge from this property registration data. First, the March 18 refinement to the program did completely eliminate the problem of invalid numbers, although recorded property transactions dropped by 75.24 percent during the next week and did not recover for the next two weeks because of technical problems with PTCL. Second, duplicate numbers (which exclude legitimate multiple sales) remain constant around 40 percent after the policy change, no matter how many transactions occur, suggesting a very regular pattern of providing duplicate numbers. Inspection of the data reveal that duplicates are commonly entered on the same day. Third, it appears that almost all of the movement in duplicate numbers prior to the program is accounted

for by duplicate numbers. Last, and importantly, there is limited evidence that officials replaced invalid numbers with valid duplicate numbers in response to the March 18 refinement of the program. In other words, the March 18 change appears to have placed a substantial additional constraint on officials' ability to undermine the program, but they were able to respond by increasing duplicates numbers to some degree. This suggests a promising direction for the program—careful refinement and calibration can remove vulnerabilities and maximize the effectiveness of the program.

Figure 2 plots invalid and duplicate numbers by district. We restrict the data to districts that have at least 12 weeks of data. Faisalabad and Sheikhupura follow the aggregate pattern, but with a larger number of spurious entries. In these cases, the recording change led to a complete reduction in invalid numbers, and created a short term drop in duplicate numbers, though this later recovered. Another remarkable feature is that in many districts, the share of weekly transactions that are duplicate numbers is around 40 percent and is remarkably stable from week to week. Of the 13 districts with 12 weeks of data, only Sahiwal has weekly averages of duplicate numbers substantially below 40 percent. Again, we see that the refinement did appear to have some effect in bringing down the total number of duplicates.

Figure 3 depicts the percent of weekly transactions that involve a number that appears at least 3 times in the registry. While in Figure 1, we have excluded duplicate numbers that belong to the same individual, by focusing on numbers that appear at least 3 times, we focus on numbers that are highly likely to not genuinely belong to the party engaging in the reported transaction. As before, the number of duplicates drops after the system no longer accepted clear false numbers, but after a few weeks delay there is a return to the normal pattern.

Evidence from Health Call Feedback

The call disposition of a sample of calls made to health beneficiaries from March 29 – June 06, 2011, depicted in Figure 4, similarly depicts some problems with contacting beneficiaries. We do not know the natural rates of invalid numbers, numbers being switched off, and the other disposition categories in this population. Contact rates are, however, somewhat low. 39.25 percent of calls placed successfully connect, perhaps reflecting problems discussed in the previous section.

We view the data on reported medicine availability, depicted in Figure 4, as potentially extremely useful to policymakers. The data from connected calls show that 37.5 percent of individuals are only getting some medicine and 28.7 percent are getting no medicine whatsoever. The online dashboard allows these

data to be readily disaggregated to the facility level. Practitioners and policymakers we engaged as part of this review were uniformly excited by this product.

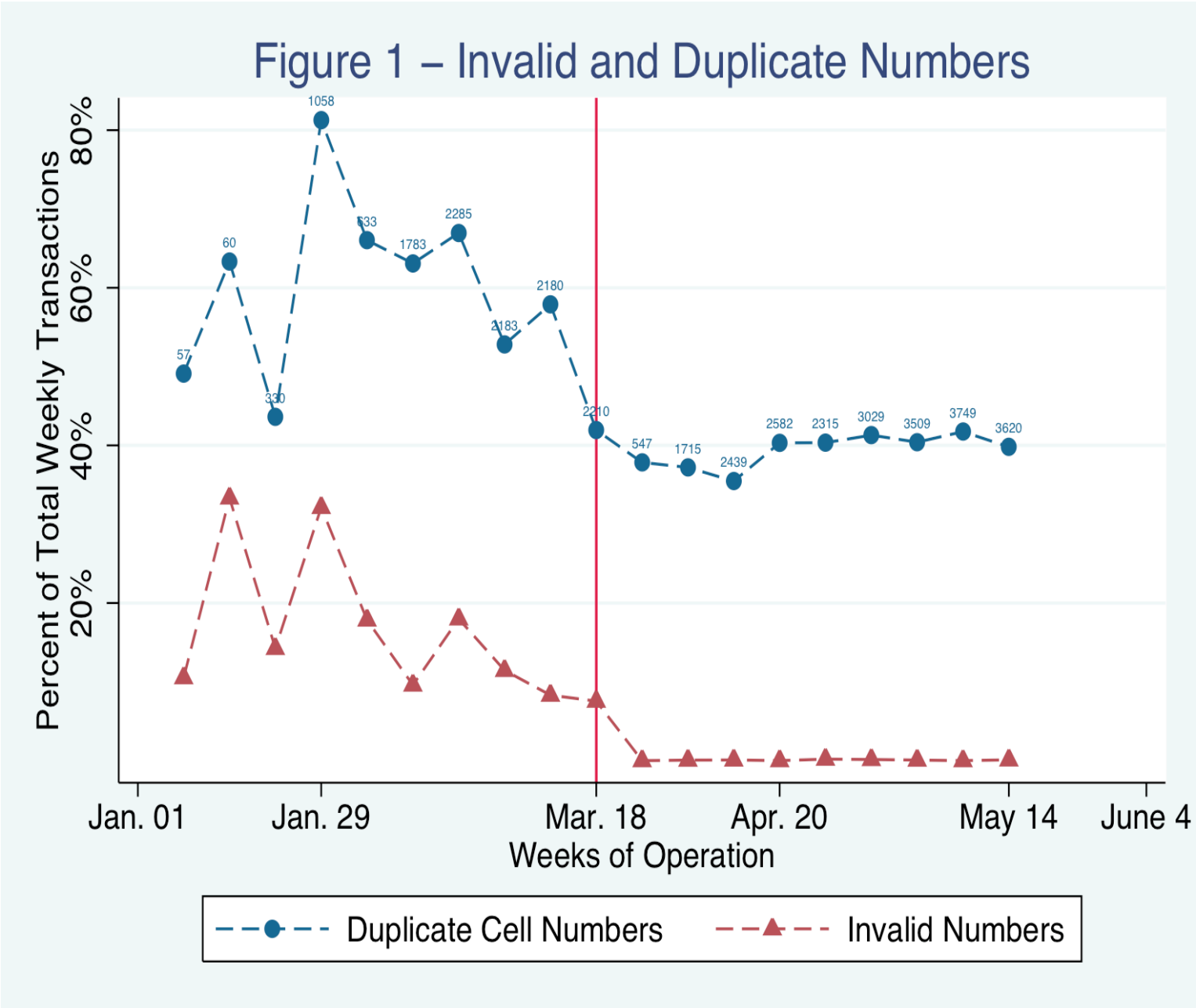
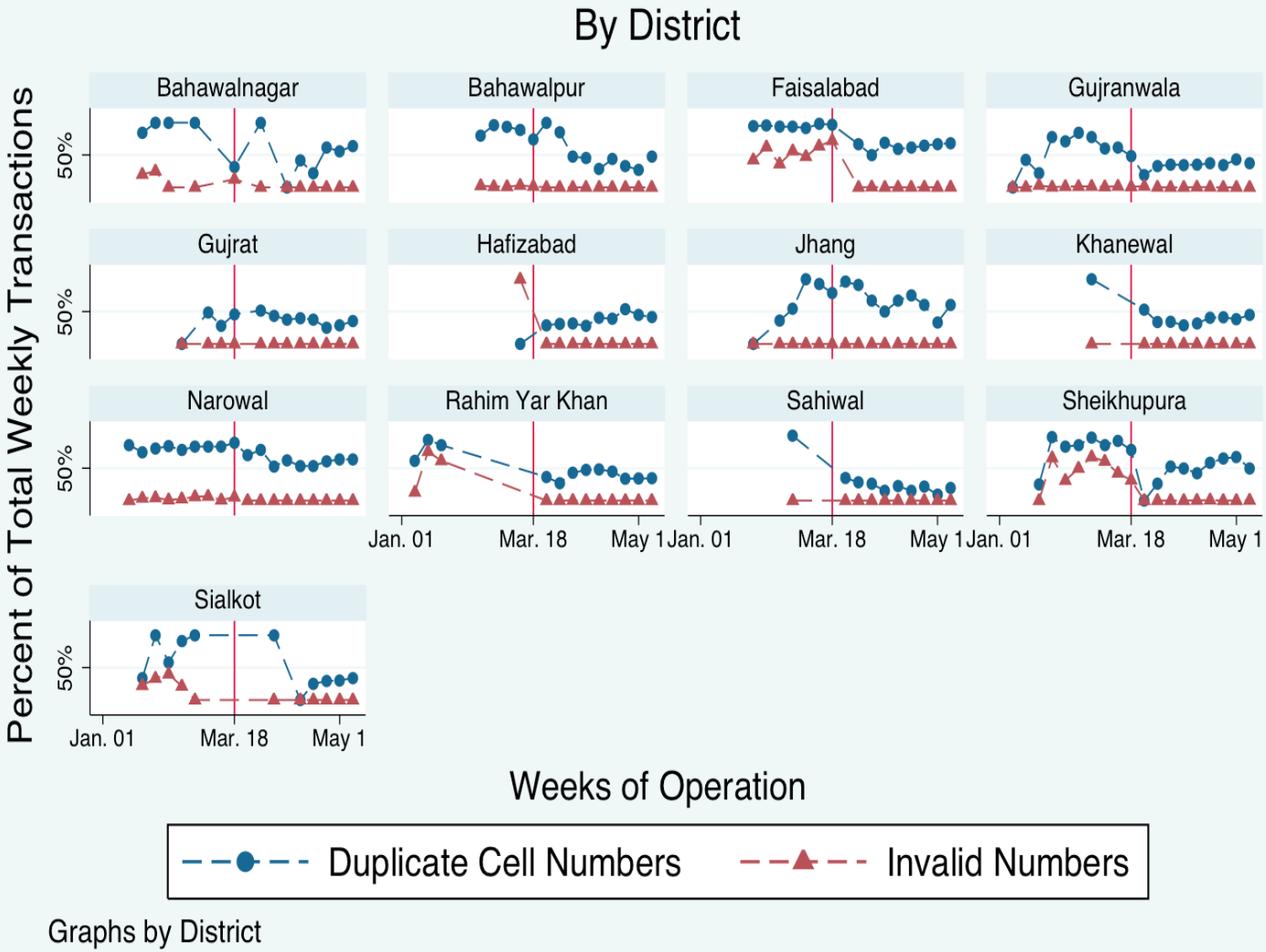


Figure 1 – Invalid and Duplicate Numbers
Percent of Total Weekly Transactions
80%
60%
40%
20%
57
60
330
1058
533
1783
2285
2183
2180
2210
547
1715
2439
2582
2315
3029
3509
3749
3620
Jan. 01
Jan. 29
Mar. 18
Apr. 20
May 14
June 4
Weeks of Operation
Duplicate Cell Numbers
Invalid Numbers

Figure 2 – Invalid and Duplicate Numbers



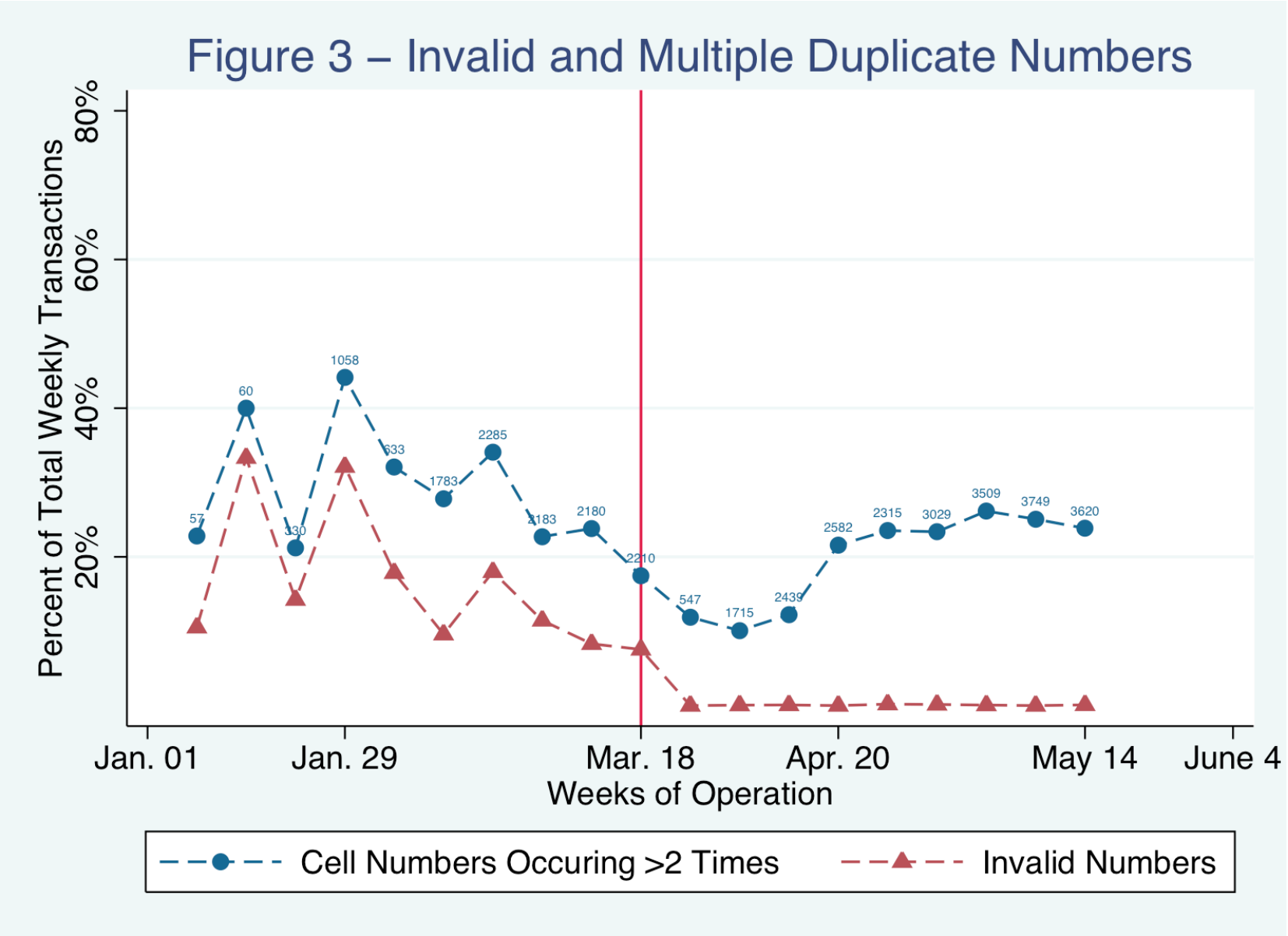
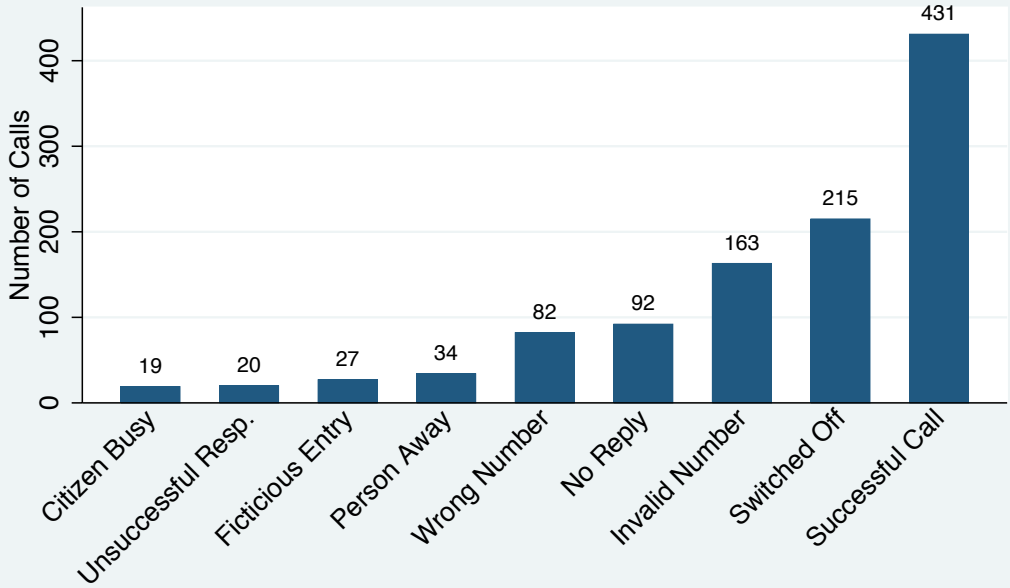


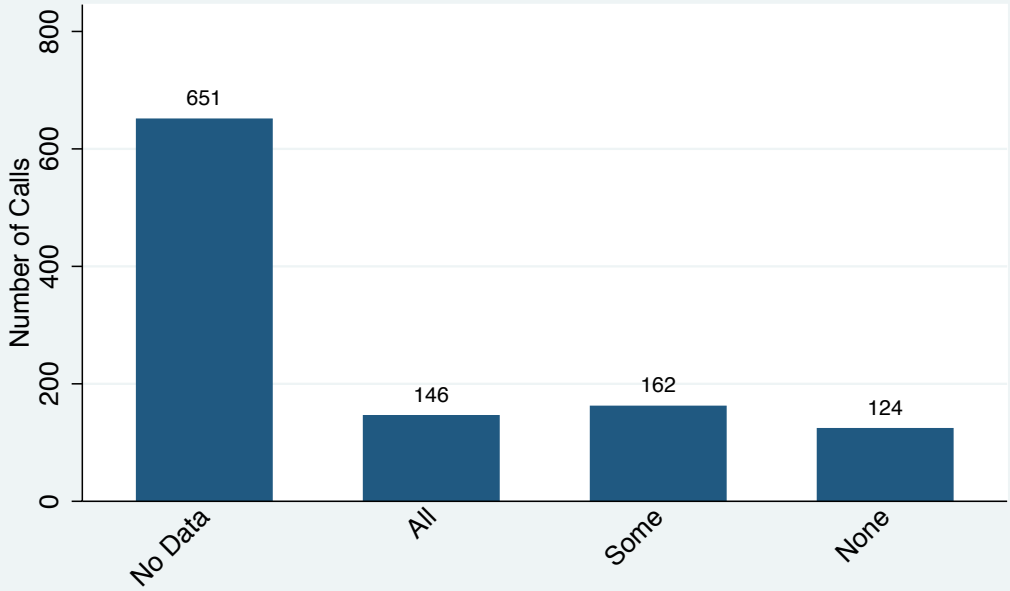
Figure 3 – Invalid and Multiple Duplicate Numbers
Percent of Total Weekly Transactions
80%
60%
40%
20%
57
60
330
1058
633
1783
2285
3183
2180
2210
547
1715
2439
2582
2315
3029
3509
3749
3620
Jan. 01
Jan. 29
Mar. 18
Apr. 20
May 14
June 4
Weeks of Operation
Cell Numbers Occuring >2 Times
Invalid Numbers

Figure 4 – Call Dispositions in the Health Sector



Note: Results reflect 1098 calls placed from March 29 – June 06

Figure 5 – Citizens Reporting Receiving Medicine



Note: Results reflect 1098 calls from 05/29 – 06/06. 447 calls were successful.

Evidence from SMS Feedback

The quantitative data above suggests that a large proportion of the numbers being provided are spurious. A brief check of the data suggests that the number provided is often that of the paralegal agent assisting the citizen in conducting the transaction (or assisting the official in collecting graft).

If our interpretation of Figure 2 is correct (that almost half of all numbers recorded in many districts are regularly duplicated and presumably belong to an official or agent attempting to subvert the process), then the feedback received from SMS or call responses will be heavily biased away from detecting corruption.

On the other hand, such biases notwithstanding, much of the feedback collected must be legitimate. In particular for offices where the volume of transactions is very large on a given day, the ability to enter valid but spurious numbers is likely to be limited to enlisting the collusion of the paralegal agents at the time the latter prepare documents. Therefore, once a few basic validation processes are incorporated, the Punjab Model retains its promise as a corruption detection mechanism.

These concerns have important implications for the nature and extent of analysis of qualitative data. In the suspected presence of spurious data, detecting quantitative results from the responses retains limited value, and the emphasis will be on developing an initial taxonomy of feedback.

For a small sample of SMS replies, we manually encoded the feedback into dummy variables in four dimensions: a) whether appreciation was shown by the citizen for the PM initiative, b) whether the experience was reported as positive, c) whether the experience was reported as negative, and d) whether the citizen seemed unclear or confused regarding the type of feedback solicited.

SMS data for 3 different dates was encoded: May 2nd (123 entries for Property SMSs and all 18 available entries on health), May 15th (204 online entries for Property), and June 21st, 2011 (all 25 available entries for driving license issuance). The total of 370 entries encoded was pooled. Given the paucity of data, and the fact that the encoding has not been validated, no quantitative implications can be drawn. However, this exercise was helpful in characterizing the broad features of qualitative responses.

Euphoria and Optimism

82 out of 370 respondents were explicitly thankful that the government had undertaken this initiative:

a.a.

Kushi hoi k hukmat jag rahi hai. Hm ne sirf srkari fee hi ada ki hai. Aur kisi kisam k koi msla darpaish nai aya.

(Hello

Glad to see the government waking up. I have only paid official fees. And faced no problem of any type.)

The idea of a responsive government seemed to unlock a sense of patriotism, which was reflected in euphoric responses such as:

Assalam O Alikum,

Mere Aziz hamwatan, mujhe koi Shikwa nahi kisi bi Ehlkar se.

Aap ke Taawon ka bohat Shukariya.

Allaha Haafiz.

(Hello

My dear countryman, I have no complaint against any official.

Many thanks for your cooperation.

Goodbye)

The gratitude was also directed towards the political party in power:

Very well Shahbaz sab i am very thankful to govt of punjab nd may Allaha bless u nd u ar company.thanks a lot

Finally, some expressions of gratitude were limited to a thank you:

No I've faced no problem. Thanks for your concern.

Reports of a Positive Experience

A total of 161 reported a positive experience at the office being investigated. As expected, almost all those who were appreciative of the effort also reported a pleasant experience (although a few exceptions are noted above).

Some respondents provided some details of the transaction:

> *Sir ji hamko koi prob ni hoi registry krvaty hoy bs actual pyment he di he jiski rcipts b mili hain thanks for Shahbaz Shareef*
>
> *(Sir, we faced no problems in getting registry done. Only the actual payment was charged, and a receipt provided. Thanks for (sic) Shahbaz Sharif*

> *Jinab mr DCO sahib st.) B.H.C.Begowala Teh.Sumbrial ne humare sath bara acha alag kya hai sehatyab ho gy tmam adoyat injection drips muft muyasar hoi. Koi paisa wasol nai kya. Amale ne bra tawan kya hai specail doctor Abdur razaq ahmad sahib incharg B.H.C begowala bre kabil mehnti naram mazaj aur ba himat doctor hai.*
> *From saleem akhtar zoja M. Tufail*
> *Vilage chak bhada teh samerial*
>
> *(Respected Mr DCO, Basic Health Unit Begowala Teh Sumbrial provided very good treatment, and health recovered well. All medicines, injections and drips were provided free of charge. Not a penny was accepted. The staff cooperated very well, especially Dr. Abdur Razzaq Ahmad (in-charge of the Unit) is learned, hard-working, kind and brave.*
> *From saleem akhtar zoja M. Tufail*
> *Vilage chak bhada teh samerial)*

Others were more general:

> *Crupption kafi had tak kam hoi hy*
>
> *(Corruption has been significantly reduced)*

However, it is unclear whether the positivity reported was due to the feel-good factor of being asked for feedback, any substantive effect of the Punjab Model on reducing corruption or improving service delivery, or due to the other service improvement efforts of the government. For example, the responses below refer to the fact that property registration now requires the verification of identity through a computerized national identity record:

> *sub kuch he thek tha photo sation aur computer ka system bauht acha hai*
>
> *(Everything was fine. The photo station and computerized system are great.)*

> *The whole process of registration of property is trarsperant and good, verification of nic to avoid any forgery @ fraud is good effort by your office*

Such responses underline the flexibility of the Feedback Model. This can be both a strength or a weakness: if less structured responses are accepted, and sieved out of the data, government officials can get data that is at least suggestive regarding how other programs are progressing. If the lack of structure distracts from the main task, however, it can be liability.

Which of these is true will depend on the context, as well as how data is encoded and analyzed.

Despite the concerns about the authenticity of some responses, there were some messages that seemed sincerely honest, and improved our confidence that with a few modifications, the Punjab Model can tap into a general willingness among citizens to help improve accountability:

> *From yasir arfat. Men ne laisence banwaya he lekin is daoran kisi ahalkar ne na to pese tlab kie na tang kia.twaqo ke khilaf acha ruwaya tha .hata ke try ke bad kafi garmi ki waja se men ne botal afar ki lekin wo log ye lene ke lie bi tyar nhen the*
>
> *(From Yasir Arafat. I got my (driving) license issued, but during this process no official solicited bribes or made the process difficult. Contrary to expectations, the behavior was excellent, to the extent that after my practical test, due to the intense heat, I offered them a cold soda and they refused even this.)*

Little Negative Feedback, and Almost no Direct Reports of Corruption

There were merely 12 messages with negative feedback. If this were representative, it would imply that merely 4% of all transactions in the offices studied are subject to corruption or poor service delivery. The anecdotal evidence is overwhelmingly against such an implication, and the result suggests that there is some stage at which the process is not yet successful in eliciting truth from a large section of respondents.. This may be, for example, due to the limited technical resources available during the current limited testing stage (the absence of a four digit short code, for example and the reliance instead on a private number). There is also no publicity and local awareness campaigns for the feedback mechanism yet. There is therefore reason to be cautiously optimistic that the roll-out of operational improvements, such as the use of a short code, mail-merge etc. will increase the response rate, and that citizens will be less guarded in their replies.

The 12 messages coded as negative responses include mild criticism of service delivery:

> *Bs rawya normal he hy yahn ka koi time tabel nhn hy per yahn ka bs guzara hy ?*
>
> *(The attitude is just okay. There is no time table. Things are merely acceptable.)*

> *Ye sab puchnay ka shukrya. Ham say salook to acha kia gia magar kisi kisam ki dawai hamko nahen di gaie sab kuch likh kar dia keh bahir say lekar aao drip+tikay +tab*
>
> *(Thank you for asking all this. They treated us fine, but no medicines were provided. They provided us only prescriptions for drips, injections and tablets to fill out privately)*
>
> *Dear sir*
>
> *All is ok but time is very vest (sic: is wasted) during registrar beyan please check it*

This wastage of time would be consistent with the theory of endogenous red tape, but corruption was hardly ever reported.

Where corruption was reported in Property Registration, it was often the usual suspect, the Patwari (a village level Revenue officer who is often sole guardian of land records, and unsurprisingly notorious for corruption) who was named:

> *Bhai.jan.aadaab.man.nay.teh.khanpur.main.10.marly.plot..ka.intqal.kraya.hai..jis.ki.maliat.160 000.hai.patvari.nay.16000.lai.hain.kia.ye.thek.hai.*
>
> *(Hello dear brother/friend. I got a 10 marla plot transferred in my name in Tehsil Khanpur, which was worth Rs. 160,000. The Patwari took Rs. 16,000. Is this acceptable?)*

> *Mutrum.mary.bhai.nay.10.marly.intqal....lia.hai.patwari.nay...amunt.zyada.lia.hain.*
>
> *(Respected Sir. My brother got a 10 marla plot transferred in his name and the Patwari took more money than acceptable.)*

The direction of these complaints is interesting: the patwari is notoriously powerful, and uniformly feared in rural settings. Why then are people willing to report his corruption, but possibly not that of the registry office?

At least two explanations come to mind: first, it could be that the feedback is largely accurate, and it is merely that patwaris remain corrupt while registry offices have been cleaned of corruption. Alternatively, it may be that the registry officers are also often corrupt, but their closeness to the transaction allow them to subvert the feedback, either through the co-opting of phone number entry (see previous section), or through a latent threat of retribution. The fact that the data may conceivably indicate two entirely different underlying dynamics is why there is such a pressing need for a rigorous evaluation that collects independent data on corruption in the same settings and is then cross-checked with data generated through the PM.

In other words, reports of corruption in the office under investigation were almost non-existent:

> *1 jaly numbrdar AKRIM NAMI 200 rupees leta ha hr 1 se uski b chuty krwa do plz*
>
> *(A fake official named Akrim takes Rs. 200 from everyone. Please fire him.)*

> *hmari ragistri Rs.1730 main ho gai ha.ham ny .Rs10000 arzi navees ko dey to us ny .Rs173 0 dy kar hmara kam krvaya baki rkam apny pas rakhi aor 2 maheny hamin tang bhi kia arzi navees ka nam abbas joya hay hamy sarkari amlay say koi shekat nahi*
>
> *(Our registry cost Rs. 1730, but the paralegal charged us Rs. 10,000 for getting this task worth Rs. 1730 done, and kept the rest himself, and bothered us for 2 months. His name is Abbas Joya. We have no complaint against government officials.)*

There were some cases of general complaints, which were often so negative that they strongly contradicted the general positivity in the data:

Har taraf he corruption hai.

(Corruption is ubiquitous.)

Significant Noise

Finally, a significant proportion of responses could not be classified for various reasons.

One problem was that the expenses for each transaction may have been different. Response such as:

37500 khrcha huwa hay

(The cost was Rs. 37,500)

Sir Ham se office wallon na 100/RS WASOOL Kiay hain

(Sir, the officials charged us Rs. 100)

mean nothing without knowing what the officially sanctioned cost of the transaction is. If the model design includes a follow-up call, such information may be useful for an agent calling back. Recently, the implementers have tweaked the data collection so that data is collected from officials regarding what was paid, and cross-checked with citizens' recollection of that amount. Small refinements like this to the Punjab Model could significantly improve both the data collected and its use.

Many seemed puzzled, not able to understand that the call was a government sanctioned one, or seemed suspicious of fraud:

?????? ???

Ap kon?

(Who are you?)

Another response, which may help begin to explain why corruption was detected at such low levels, is that others seemed to worry that the investigation was of them, not the officer, and were defensive in their responses:

Muhtram. mainay government k wajbat k illava koi passay nahi diay.

Sir, I have not paid anything beyond official expenses.

Asking citizens whether a bribe was taken from them is equivalent to asking whether they committed the illegal act of paying a bribe. Even if it is muted by a culture of corruption, there may be natural hesitancy in truthful reporting in such a case.

Another aspect of inquiring whether a bribe was paid is that people may be offended, as this next person seemed to be:

> *Nhi muj sa kisi police waly na paisy nhi liya or jo driving licence ma na hasal kiya ha wo apni mahnat sa trafic signs or drving test pass kar k liya ha ok*
>
> *(No, no police officer took any money from me, and the driving license I have gotten, I got by putting in effort to pass traffic signs and driving tests, okay?)*

As expected with a newly implemented program in a country with poor education standards, there was a significant number of unclear answers that betrayed either a failure to understand what information was requested, or replied with information beyond the scope of the program, sometimes tragically so:

> *muhtram janab DCO sahib aap ka sms recived kia dil main umid ki kirn photi actually 4 /6/2006 ko mera chota bhai kidnape howa tha laikan abhi tak mil nahi saka hamara muqadma no 442/2006 hai jo settelitetown police station main registered hai aap se request hai k aap is case k leye koi special team appoint karain k mera bhai bazyab ho saky aap k jawab ka muntezir Taqi abbas dar*
>
> *(Respected DCO, receiving your message has lit a lamp of hope in my heart. Actually, on 4/6/2006, my younger brother was kidnapped but has not been found yet. Our case number is 442/2006, which was registered at the Satellite Town Police Station. I beseech you to appoint a special team for this case so my brother can be recovered. Awaiting your reply, Taqi Abbas Dar.)*

To summarize, a very preliminary qualitative study of the data suggests that there is great optimism and positivity associated with the mere fact that, for perhaps the first time in many citizens' lives, the state is actively soliciting feedback. The very fact of eliciting responses is a massive factor in favor of the program, and consistent with the expectations of those running it.

Unexpectedly for a country with weak governance, and where corruption looms large in the public imagination, actual reports of corruption are almost non-existent, and reports of a favorable experience are very numerous. This is a surprising finding that merits further investigation. There are many possible explanations: the offices where the program has been implemented may be associated with the DCOs most proactive in making sure it is implemented, and this may have increased their surveillance of the implementing offices in other ways also. It could be that other reforms, such as cross-checking property registration applications with a centralized identification system have translated into actual and substantive improvements

on the ground. Or it could be that paralegal agents and officials colluding in a corruption racket may have effectively subverted the transmission of the citizens' feedback. Perhaps the truth is a mix of these and other factors.

The preliminary findings above are made even more surprising by the fact that senior bureaucrats have found the project to be very worthwhile. In June 2011, the Commissioner of Bahawalpur took disciplinary action against nine government officials on the basis of complaints received in PM reports by the DCO of Hafizabad. Moreover, other senior level bureaucrats and politicians have also reported a very positive experience with PM, among them many DCOs. These officials value that the project sends them a list of random citizens who had transacted in the various offices under their control recently. Often, these officials report favorably on the experience of being able to talk directly to the citizenry, and also report that the citizens were often deeply moved and provided feedback openly.

There are plausible explanations for why the reports being received are so positive: the most optimistic is that concurrent improvements in service delivery mean that people actually have less negative feedback. Another is that they feel distant or fearful of the official, and report positively despite the repeated and sincere advice of the official to speak freely. Instead of speculating on the truth, the implication of this puzzlement is the need for a rigorous verification of the PM process, through a comparison with another feedback mechanism (such as surveys).

The implementers too have struggled with the question of why there are so few reports of corruption. They have recently changed the questions being asked, and tried to identify key words from responses that might articulate why people are reporting significantly less corruption. However, this exercise has not yet been studied.

The Punjab Model has made an important step in initiating the effort to elicit citizen feedback, and has brought down many cultural and historic barriers in this regard. However, as a young program, it has many easily achieved improvements possible, and some of these are discussed in the next section.

1. Issues and Recommendations

Introduction

The Punjab Model program uses existing technologies in an innovative and unprecedented way. This necessarily implies that there is much to be learned, and that there is likely great scope to refine and develop the intuitively appealing underlying idea that increasing feedback from citizens who are provided a service by an official can improve governance.

If successful, the model has great scaling potential, across sectors, provinces, and ultimately countries. The manpower and capital outlay required to implement the program are minimal, given that it exploits low cost ICT technologies and call centers, which are particularly cheap in South Asia, and is suited to viral adoption. For these reasons, the program is potentially extremely cost-effective.

This section discusses issues that can arise, possible limitations of the design, and our recommendations on how the program should move forward.

Structuring an Evaluation

While the Punjab Model designers emphasize two main elements: citizen engagement and citizen feedback, our evaluation is primarily of the Model as a feedback-collecting mechanism. This is not because of an undervaluing on our part of the first element; we report the strong optimism and gratitude shown by citizens when they are contacted in an earlier section. However, we believe that the positivity associated with merely being contacted is ultimately founded in the expectation that the feedback provided will matter, at least broadly. Therefore, we expect the future viability of the Model to depend greatly on the extent to which it is successful in detecting corruption, and on the extent to which officials use the information provided productively. As a result, we consider it sufficient to limit the discussion in this section to a preliminary evaluation of the Punjab Model as a feedback mechanism.

A full evaluation of the Punjab Model should, we believe, be framed as a Cost-Benefit Analysis. Does the output justify the input of time, energy, and public monies spent? The Punjab Government has limited resources to spend on many different interventions. Is spending the last penny of public money on the Punjab model more or less effective then spending that last penny on, for example, improving ambulance coverage, or providing subsidized housing? Is allocating a extra official to work on the Punjab Model more useful than allocating him to facilitating coordination between different offices within the district? To the best of our knowledge, no such calculus has been undertaken for this or any other public project either in average terms or more importantly, on the margin. This void needs to be filled in as quickly as practicable, preferably for all major government projects. We emphasize that the costs and benefits outlined here are based on limited data and so necessarily somewhat speculative. We do, however, advocate a thorough cost-benefit of the program according to international best practice. Our team intends to pursue this in the coming months.

Cost Analysis

We believe that the core costs of the program are the direct financial costs, with important caveats emerging only if a) reporting corruption causes retribution against citizens by officials, b) PM slows down or otherwise has a deteriorating effect on service delivery, and c) if collecting and reporting this information taxes the attention of government officials (in particular the DCO) to the detriment of their other tasks). These concerns bear careful monitoring, but have not been observed during our preliminary assessment. Therefore, our current expectation is that accounting costs are a reasonable first proxy of actual economic costs of this mechanism. We focus our attention in the rest of the section on considering benefits and suggesting ways to increase these.

Benefit Analysis

In evaluating the Punjab Model's effectiveness, we believe that three broad questions need answering to evaluate the program and its usefulness. They are best posed in the following order: first, is the information collected by the mechanism accurate? Second, is the right information being fed back to the right people, at the right times? Third, does the Punjab Model as a whole have an observable and salient effect?

(a) Evaluating and Improving Information Generated: Is the Mechanism generating data accurately?

As reported earlier, the number of responses reflecting dissatisfaction is unexpectedly low. There are a number of reasons why this may be happening, and understanding these is fundamental to the future viability of the program. The objective of the ongoing research and design refinement should be to get the technology to the point where it wins out, and largely precludes data scamming and also captures accurate and representative data on corruption.

We support the decision made by the program designers to first target corruption, which is a narrow and precise concept, before expanding to issues of general satisfaction with service delivery. More data on corruption, both precisely elicited via SMS and phone calls, and from separate measurements to corroborate and calibrate the quality of the data coming in via internal channels, are critical for two reasons.

First, precise data are critical to the operation of program. While fudges may provide proxies for the extent of corruption, the core logic of the program is that it closes information gaps, which all malfeasance needs to thrive. Maximizing the fidelity and accuracy of data generated by the model will improve its function and its deterrent effect. Correspondingly, suitable protocols for organizing, storing and transmitting the data need could be improved.

There may be two ways in which inaccuracy is entering the data collection: it may be that citizens are not reporting accurately, or it may be that officials are not providing accurate numbers, as discussed above.

(i) Issues with Citizens' Reporting and Recommended Fixes

It is instructive, when thinking through the problem of inaccurate feedback by citizens, to consider the incentives they face in reporting the truth or not. When the citizen weighs the possible negative consequences against the positive effects of reporting corruption truthfully, it may be that the former outweigh the latter. If they fear negative consequences from the official, or a racket providing patronage to that official, in the event that they report the truth about the existence of corruption, even if the fear is unfounded, they are naturally less likely to report it. This problem may have particular bite in the case of corruption, as the same officials who are asking for bribes are recording numbers.

Citizens are also unlikely to report if they fear that the information could be used for other purposes. This is especially salient in property registration and other tax-related transactions where auditing is a concern. This issue is compounded by the heightened sensitivities over providing numbers given the current security climate and associated stringent control of cellular numbers in Pakistan. Last, individuals need to know that their responses are held in confidence and will result in some benefit either to themselves or their community. Whistle-blowing would have few positive consequences besides greater self-worth even if the information was being acted on with the full weight of the government behind fixing each office. In a world where negative reports are likely from each office, and citizens have grown up without expectation of government action even with overwhelming evidence in hand, these benefits of whistle-blowing go away, and as with all public goods, whistle-blowing is underprovided.

Uncertainty about the origin of the call is likely to stifle truthful reporting: one is much less likely to report corruption to someone whose identity is suspect. The current efforts underway to mask the source number for SMS and to have calls and SMS messages originate from a well-publicized short-code will surely help. This will eliminate confusion about the source of the call, thereby removing one contributor to underreporting. Related publicity about how this information is being used shall also help, as would the establishment of a helpline where anyone could call in advance and get information about the Punjab Model and ask questions. This could reinforce the idea of providing feedback in citizens. A suggestion from our call center focus group was to make sure that citizens whose feedback led directly to disciplining action or other service delivery improvements be sent an SMS thank you note. Such

small details may go a long way in sustaining the current goodwill attached to the program.

Better still, a feature can be built into the project to credibly signal the confidentiality of the citizen. A double-blind procedure can be adopted, where for example, two separate operators call the citizen: the first verifies that the correct person is on line, and a few minutes later, the second calls from a separate number (or better, a separate office) to ask about their experiences. A commitment would be made and advertised that no single person in the system can access individuals' comments, even when data for the office as a whole are freely available. This would also fit well with the idea of the project as an information aggregation system that specifically and empathetically states itself as not being a complaint receiving operation.

Finally, whether citizens report truthfully and accurately or not can depend on the questions asked and other scripting details. The program implementers are already experimenting with these, but there is the need to add rigor to how new scripts are generated and evaluated. In particular, as the PM is rolled out to various departments, sector-specific scripts should be evaluated, evolved, and incorporated. Focus-groups with citizens, officials and calling center agents should be regularly held during such efforts. These simple efforts can sometimes help iron out wrinkles that matter.

(ii) Issues with Officials' Transmission and Recommended Fixes

For various reasons, officials may have low incentives to report truthfully too. This could be due to misrepresentation, or due to a lack of positive incentives to report accurately.

It may be that phone numbers are being misreported: corrupt government officials or paralegal agents may be providing their own phone numbers, or those of others sympathetic to their attempts at subverting the program. When phone calls get made to these numbers, the respondents may lie and act as if they were the ones provided the service. This is a problem the call center agents had suspected. Iteratively experimenting with and picking wise filters that identify and ignore such numbers is crucial for the data generated to be valuable to its recipients. Automatic cross-checking of numbers against the universe of numbers associated with previous transactions in that office or district, in particular when the citizens' names don't match, would be an important first step. Those submitting such numbers must also be traced through monitoring and sanctioned.

Another way in which misreporting may happen is that a large fraction of phone numbers reported is clearly invalid. Some of this reflects officials misunderstanding about what needs to be entered and lax data entry effort. This can be easily fixed through increased training and assistance, and will in

any case diminish with time as more and more officials become familiar with the paperwork. Some portion of this problem likely represents incompetence in entering data accurately. However, a sizable proportion of these invalid phone number entries is clearly a deliberate attempt to undermine a program of corruption detection. This may be best tackled if a clear and intelligent protocol for pattern recognition is built into the system to detect which offices the invalid numbers come from. A protocol should then be built in, whereby offices submitting invalid entries regularly are investigated. Another effective approach might be to delink the collection of numbers from the point of transaction. This could be achieved, for example, by providing citizens with a short code and asking them to text information according to a specified format. The short code could be provided in a pamphlet instructing that they should vigilantly protect their identifying information from the recording bureaucrat. Alternatively, a different official not responsible for the transaction could record the number.

Also, the program has not yet allocated a budget, or official phone for officials to use in sending reports. The lower to mid tier staff implementing the program have often dipped into their own earnings, or reallocated money earmarked for other projects to send in the messages daily. While SMS bundles are quite cheap, asking officials to incur a cost without providing them an associated benefit may potentially discourage them from sending in complete data daily, by violating their sense of entitlement or fairness. If this feeling is widespread, mid-level officials in tacit agreement may even look the other way, undermining the program.

A central feature of the Punjab Model is that the officer conducting the transaction (and against whom the feedback would go) is the same one soliciting and entering the data. In some cases, where the official conducts possibly a hundred or more transactions per day, his ability to manipulate data may potentially be limited (although this is far from certain, especially in the suspected presence of colluding paralegal agents). However, in other cases with a low number of transactions per officer and data entry being done far from the oversight of superior officers, it may be a much larger problem.

This data management by the officer is a feature of the current operation that is practically necessary, but ultimately less desirable. As a general rule, it is potentially easy for them to provide invalid numbers, or enter numbers where the call recipient will pretend to be the citizen, but will be someone sympathetic to the corrupt official. As mentioned above, this ability will be constrained if the number of transactions is high, especially with the expected introduction of smart filters that check entered numbers for validity and repetition, but will be a limitation for contexts of fewer transactions, as discussed above.

We provide three recommendations for fixing this. A specific actionable item we recommend would be to record the answer to a "security question" such as those asked in order to provide internet users with forgotten passwords to online accounts. These questions could be a citizen's father's name or some other item known uniquely and easily recalled by the citizen. This would add a layer of protection against affiliates of corrupt officials impersonating service beneficiaries, which is currently happening at least on some scale.

Second, monitored officials could be provided with a dedicated phone which, for every transaction, requires that the citizen's number be immediately texted in. In response to the official's text, a unique serial number required to complete the transaction could be provided in a return text. This provides the service beneficiary with strong incentives and some control to guarantee that their number is entered correctly.

In the longer term, the ultimate goal of the program (although this may not be possible in the medium term) should be that eventually and when cellular phone records are updated and sufficiently verified to reflect actual user data, the collection of cell phone numbers should be replaced, and CNICs only should be accepted. Then, the Cell Phone Registry can be used to pull the phone number corresponding to that CNIC and a call can be placed to that phone number. There are many benefits to this: first, officers are far less likely to wrongly enter CNIC data, or leave the field blank, and citizens will likely insist firmly that CNIC data be entered correctly, because this can void the transaction itself. Second, it preserves the anonymity of the citizen far more. Names are far easier to remember than numbers, and neither officers nor calling center employees will be able to remember CNICs to identify citizens maliciously. This recommendation is clearly not realistic at this stage: the cell phone registry is not yet sufficiently accurate, and it is common for phones to be bought against a CNIC not belonging to the end user. However, it is important for there to be agreement of this as a desirable end goal, to be targeted when the Cell Phone Registry has been improved sufficiently.

The implementers are already considering this. NADRA is the National Database and Registration Authority charged with the establishment of a new registration system for the entire population of Pakistan. It offers CNIC verification services in a pay-per-use model. PTA is the Pakistan Telecommunication Authority that, among other things, maintains a database of cellular connection owners. The program designers propose that a service be developed that would let the system poll NADRA's database for the person's CNIC record and then poll PTA's database and get all cell numbers registered against that CNIC. This would effectively eliminate the need for identifying information beyond NIC number (since the system would pull numbers from

PTA, and citizens won't to divulge it directly), and let the government track multiple interactions done with the same people helping identify usage patterns.

Finally, a problem that remains despite all of the possible improvements listed above is one that potentially positively biases the information collected. This mechanism collects data at the time of the completion of transaction. Ideally, feedback should be solicited from the entire population of those who visit a government office on a given day. Instead, the feedback is being gathered only from those who successfully completed the process. Transaction completion is presumably quickened by the payment of a bribe, and it could be that when this happens, the citizen chooses to provide positive feedback in the fear of being implicated for his payment of a bribe. Thus it may be that a citizen completing a transaction and providing feedback is more likely have paid a bribe and then reported untruthfully. To the extent that this problem is salient, a traditional complaint service is superior (although PM has many features that are in turn superior to a complaint service).

(iii) Evaluating the Information

A key problem evaluators will face is determining whether the information being generated by the system is accurate or not. In our view, the long-term success of the program hinges to a large extent on this. While some traction can be found on this problem by considering proxies such as invalid number entry, these proxies are problematic because a) they provide at best an estimate of corruption that can be defended as reasonable, but that has not been shown empirically to be correlated with corruption, and b) because by their very nature, the proxies themselves are worth eliminating from the system and will become harder and harder to detect. This has already happened to duplicate number entry, and will soon happen to repeated number entry too.

There is a need therefore, to test the accuracy of the information generated in the system by collecting data on corruption through other ways as well, and then comparing the data generated by these different measures. We will come back to ideas about evaluating in subsection (c) below.

(b) Evaluating the Use of Information: Is the right information being fed to the right people at the right times?

The question of how feedback is organized into reports and fed back into the system is also one needing careful consideration. To date, it is left to each recipient of the data to decide how to use it.

For the Punjab Model to be especially useful as it scales, the data generated must be meaningfully analyzed, summarized, and provided to those in the chain of command who have the right incentives, and the legal jurisdiction to take corrective action where necessary.

(i)Data Analysis and Presentation

Ideally, feedback in the future will be coded at the time of receipt: SMS feedback could benefit from automated content analysis, which would scan each incoming SMS and detect different characteristics (for example, whether the words "thanks", "CM", "corruption", or "Rupees" occur. Operators making phone calls for feedback can be tasked with rating the call on 1-5 scales while the call is ongoing in a few basic ways (degree of appreciation shown, degree of understanding or confusion shown about the system, corruption reported, quality of service indicated, etc.). The accuracy of the encoding can be checked periodically and improved (automated content analysis algorithms may be improved, and human operators may be rewarded or punished based on this accuracy).

An intelligent alternative mechanism of coding responses that the program implementers are currently testing is to crowd source the coding of responses, by allowing people to log in to a system and code each text response as negative or positive, reporting corruption or not, etc. This is a very appealing initiative, and if scaled with proper security, redundancy (making sure each statement is coded by 2-3 evaluators) and publicity, could be a very successful alternative to content analysis, and may help build the sense of community by allowing citizens to engage proactively in corruption detection efforts.

It is also very important to ensure that data is available in real-time, and there would be great value to ensuring that the presentation of the data is done in an engaging way. The program implementers should take on the lofty aim of producing a data engagement interface so appealing that senior officials will consider it a joy to log in and compare different metrics available to them.

(ii) Data Availability and Use

An additional question that needs further attention is the problem of who gets access to the information generated at a particular office (or alternatively, what information is available to each officer with access).

It is instructive in developing an answer, to refer to the problem discussed earlier of dishonest reports against an honest officer, and the current modus operandi of allowing each district and each officer to reach its own decision regarding how a case of negative reports is to be handled.

There are clear problems with adopting the far more mechanistic alternative of treating the citizen feedback as a statistic that directly determines job outcomes and prospects: the specter of jealous or ill-meaning paralegals, rival officers, or others conspiring to provide overly negative feedback against an honest or anti-social officer cannot be ruled out. The question that then arises

is: should a maximum threshold of negative reports against an official be defined at all?

This is a rather difficult question: a DCO or EDO has jurisdiction over how to act when presented with any evidence against a subordinate. It is possible for them to use a single piece of evidence to justify punitive action, or to disregard even a preponderance of evidence. To remove this degree of control from the office of the DCO or EDO is neither desired, nor practicable. There may, for example, be evidence outside of the PM that points in the opposite direction from what is reported through the Model. It is not uncommon among bureaucratic circles to hear stories of a dynamic officer finding a novel way to detect corruption, perhaps by enlisting outsiders' help or going undercover themselves. The PM is a compliment to, not a replacement for such efforts.

The question that follows then is, if some bureaucrats can be entrepreneurial in detecting and eradicating corruption, why aren't other officers also proactively reaching out, instead of waiting for complaints to come in?

If the answer lies in a natural variance of quality, effort or commitment amongst bureaucrats, the designers of the PM can attempt to leverage the dynamism of better officers to pressurize the less dynamic ones, by providing information at different tiers of the government that reports not just the absolute corruption rate reported, but what that rate is in comparison with other districts. In-group reputation can be a powerful motivator, and knowing that his district has ranked low consistently can make a DCO look bad amongst his peers and possibly push him into action. A thoughtfully designed and attractive graph that ranks districts along different margins in real-time may be a very powerful instrument that finesses the problems of a strict threshold of negative reports that then leads to mechanistic consequences.

The basic principle in organizing information access should be that an official has access to summary reports on all offices falling under her jurisdiction, along with averages from a relevant cohort. Therefore, an EDO (Health) should be able to look at all reports from her District and Tehsil Hospitals, Rural Health Centers, Basic Health Units, and all other areas of responsibility, along with averages of reports of the same type from all other districts. On the other hand, the doctor in charge of a Basic Health Unit should only be able to access the report for his Unit, the average in the district, and the provincial average.

In various other settings, such as universities, hospitals and other businesses, employees are required to log in to various educational or organizational programs at various intervals. It may be useful to require officers who have easy access to computers to log in, say, once a month and click through different graphs highlighting the performance of the offices they are

responsible for, relative to other similar offices. An officer one grade senior to them would then be provided information about the completion of this overview, and could nudge the junior officer to complete requirements informally. Such requirements are easily circumvented (handing the password to your children or assistant is easy enough), and depend on the senior officer's interest in enforcment, but are also technically easy to implement and won't be completely without benefit.

Additionally, offices which are reported in the bottom bracket provincially (i.e. the ones with the most negative feedback), especially on questions of corruption, should be reported to the Chief Minister's Inspection Team (CMIT), the Punjab Ombudsman or the Punjab Anti-Corruption Establishment Department, as appropriate. However, care must be taken to decide that the report gets sent only to the one organization deemed most appropriate to deal with the issue, in order to remove the possibility of the classic public good problem, i.e. that knowing other inspectors are getting the same report can cause an inspector to engage in passing the buck behavior.

It is also recommended that the reports generated through the system be made publicly available on a website, perhaps with a time lag of a few weeks or a month (to protect privacy of citizens). With citizens providing feedback, the natural extension is to allow citizens to help scrutinize the data and identify patterns and areas of concern.

Finally, a key achievement of the PM is that, perhaps for the first time in Punjabi history, a real-time performance monitoring mechanism has been implemented. RHC deliveries, property registrations etc are reported to the center daily, and the system could quickly be adapted to include any other activity the government wishes to monitor. This monitoring ability does not need to be used only for corruption detection, but can be expanded into other uses to predict where, for example, critical medicines may be running short and need restocking soon, and for epidemiological forecasting.

(c) Evaluating the Punjab Model

The ultimate question any benefit analysis must ask is: what is the extent to which the Punjab Model, in any given guise, affects outcomes, and how are those impacts to be valued?

It seems natural that the PM should be evaluated principally in terms of the extent to which service delivery is improved and corruption reduced when it is implemented, compared to when it is not in use. To do this, as was mentioned earlier, it would be useful to first compare the data generated by the PM with other measures of corruption and service delivery, to evaluate the PM as a credible yardstick, and then to study whether its application is improving indicators in the areas to which it is applied.

A key method of conducting such an evaluation that should be considered is to have mystery shoppers visit a point of transaction and collect data on observables that may correlate with corruption. For example, measuring the time taken for a person standing in line to come out of the office having completed their transaction, the number of people standing in line, the number of paralegals observed in the facility, or the average bribe a paralegal solicits when approached. This information would then be compared with data collected through the PM for that facility to observe the extent of correlation. Such an evaluation is critical to any analysis of the PM data's credibility.

Another possibly less precise comparator that could be used in a limited roll out would be to provide a website or SMS number where citizens log or dial in to provide voluntary feedback about the visited office. Numerous experience goods and services are rated informally on the internet (through websites like RateMyProfessor.com, star ratings on Amazon.com etc). The degree of divergence in PM from this would probably be of less use than the mystery shopper method discussed above however, since the data gathered through such a mechanism is expected to be substantially noisy.

The next evaluation question to ask is: is PM helping improve outcomes? One way to test this is to evaluate the service delivery and level of corruption in offices where PM is rolled out, and compare to the same indicators in other, similar officers where PM is not rolled out.

Among the many attractive features of the Punjab model is that large amounts of data are generated very quickly. This means there is a favorable, built-in environment for constant study of the data, its analysis in real time, and for a quick turnaround in improving design. Unlike most public projects that a government undertakes, the Punjab model has a built-in capacity for constant data-based evolution. This capacity for constant improvement and feedback will remain untapped however, without a scientific R&D effort.

Many of the responses to these problems involve intelligently refining scripts, number collection procedures, and disseminating credible information about the purpose of the Punjab Model. We believe this to be an area that is especially promising. The data generated by the model arrive at a rate that would allow a quick assessment of the effect of different variations of the model.

There are already some spinoff efforts originating from the PM, which speak to the easy scalability of the model. There is currently a World Bank effort (which won an Innovation Fund award to reach out to school management committees) to reach out to parents of students in Sindh Province. There is also a related effort in Punjab to make sure that supervisory officers in the health sector are performing their duty and taking measurements of health assets by having them complete a geo-tagged survey of medicine and

health assets and related beneficiary data. A rigorous evaluation is being considered in these areas and PM too would benefit from the conducting of such a study.

Learning from the Analysis

As mentioned earlier, evaluating the benefit of the Punjab Model requires step-wise answers to the three questions posed above: (a) is the information collected by the mechanism accurate? (b), is the right information being fed back to the right people, at the right times? (c), does the Punjab Model as a whole have a favorable effect?

It should be expected that due to the innovative nature of the program, many problems and deficiencies will be detected. Given the intuitively appealing nature of the project, it is best to think of the problems detected, not as mortal strikes against the program, but more as wrinkles in the project's current guise, to be ironed out in future design iterations.

To date, the Punjab Model has been run as an entrepreneurial project: the focus has been on getting the project off the ground and overcoming challenges as they arise. As it scales up to become a province-wide program, it may make sense to shift from being entrepreneurial to being more systemized.

As others join in the effort to scale, we recommend that the core existing team undertake a documentation. Past challenges faced and resolved, current practices, and ideas for the future should all be distilled into a master document. This will serve two purposes: first, the project will not remain as dependent on individuals and their private knowledge. Second, such documentation will allow those with less day-to-day knowledge, but with expertise in a relevant area, to make useful suggestions. An IT professional might, for example, look at a mapping of information flows and suggest existing technologies to improve a process. A department secretary whose department was not considered appropriate to scale PM up in might figure out ways to get around perceived problems, etc.

An example of an easy fix emerging from an effort to document came from our Call Center focus group, where agents complained that they had sometimes lacked updated names of DCOs in the scripts they had used, and citizens had ridiculed them, undermining the program's credibility. This was a weakness of the previous implementation that was obvious to one group of workers, i.e. the calling agents, easy-to-fix, but perhaps not visible to the program designers at the time amidst all the other logistics that needed sorting out. As soon as the focus group helped us identify this small wrinkle, it was reported to the program implementers and a fix agreed to. Fine-tuning the processes in this manner is an easy but valuable idea worth institutionalizing.

Finally, as problems and solutions are iterated through, they will help us form a more complete picture of what enabling factors are needed in an environment to make the PM work. It makes sense to pick the most favorable transactions for initial roll-out, as PM has done. Perhaps the government should task each department with building a taxonomy of the different transactions its officials conduct in one-on-one interaction with citizens. The measurability of outcomes (to facilitate an evaluation), non-collusive rather than collusive corruption, little repetition of transactions and the lack of officials' ability to engage in retribution seem like important factors determining the success of PM in a given context. To this should be added a ranking of the importance of the transaction to citizens' welfare. Such a suitability index can then be used to guide the scale-up to other transactions in the future.

Conclusion

We have been engaged in a brief, qualitative evaluation of the Punjab Model in which the program comes across as very promising, but with significant concerns regarding whether it is currently accomplishing the task of accurately collecting information about corruption and service delivery.

The idea of a government engaging in the elicitation of citizen feedback on service delivery, and requesting evaluations of levels of corruption are deeply appealing, and therefore we are inclined to view the problems that exist as defects to be iterated out. We are cautiously optimistic that, if scaled up carefully and through the employment of scientifically rigorous design iteration and evaluations, the PM can be a useful addition to the array of Punjab Government feedback mechanisms.

However, we also strongly believe that best-practice demands that the mechanism be rigorously and fully evaluated by validating the information collected, verifying that reports of low corruption correspond with actual decreases, and by using a full, rigorous evaluation to determine whether the PM is positively affecting citizens' transactions with the state.

The Punjab Model is a very promising mechanism using existing technologies and easily replicable elsewhere. The idea of citizens providing feedback after completing a government transaction is intuitively and broadly appealing. Much progress has already been made, perhaps chiefly in overcoming natural inertia and rolling the program out at in a pilot stage in 13 districts. However, as with any innovative, path-breaking project, the need for certain refinements has already been made. The program, while very promising, would benefit immeasurably from putting in place a system that regularly and rapidly tests new implementation methodologies, scripts and the like and

occasionally observes outcomes independently of the reported feedback, to provide the Punjab Model feedback of its own.

Bibliography:

Acemoglu, Daron and Thierry Verdier, "Property Rights, Corruption and the Allocation of Talent: A General Equilibrium Approach," Economic Journal, September 1998, 108 (450), 1381 – 1403

Aker, Jenny, Paul Collier, and Pedro C. Vicente, "Is Information Power: Using Cellphones During an Election in Mozambique," Working Paper, Oxford University November 2010

Ansar Abbasi, "No magic wand, still a DCO works wonders!" The News, Saturday, July 05, 2008

Ansar Abbasi (2010) "Jhang model to cut corruption, now all over Punjab", The News, April 14, 2010

Bhatti, Zubair K. "Fighting corruption with cellphones", June 15, 2009

Becker, Gary S., and George J. Stigler, "Law Enforcement, Malfeasance, and Compensation of Enforcers," Journal of Legal Studies, January 1974, 3 (1), 1 – 18

Bertrand, Marianne, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan, "Obtaining a Driver's License in India: An Experimental Approach to Studying Corruption," The Quarterly Journal of Economics, 2007, 122 (4), 1639 – 1676

Besley, Timothy & Robin Burgess (2002) "The Political Economy Of Government Responsiveness: Theory And Evidence From India," The Quarterly Journal of Economics 117(4): 1415 – 1445

Cadot, Olivier, "Corruption as a gamble," Journal of Public Economics, July 1987, 33 (2), 223-244

Callen, Michael and James D. Long (2011) "The Political Economy of Election Fraud: Evidence from a Field Experiment in Afghanistan" UCSD Working Paper

Chattopadhyay, Raghabendra and Esther Duflo, "Women as Policy Makers: Evidence from a Nationwide Randomized Experiment in India," Econometrica, September 2004, 72 (5), 1409 – 1443

Economist Article Eureka Moments September 24, 2008

Ferraz, Claudio and Frederico Finan, "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes," The Quarterly Journal of

Economics, 2008, 123 (2), 703 – 745

Fujiwara, Thomas, "Voting Technology, Political Responsiveness, and Infant Health: Evidence from Brazil," 2010. Mimeo, University of British Columbia.

Jack, William and Tavneet Suri (2010) "The Economics of M-PESA," MIT Working Paper

Jensen, Robert (2007) "The Digital Provide: Information (Technology), Market Performance and Welfare in the South Indian Fisheries Sector," The Quarterly Journal of Economics 122(3): 870 – 924

Khwaja, Asim Ijaz and Atif Mian (2005) "Do Lenders Favor Politically Connected Firms? Rent provision in an Emerging Financial Market". The Quarterly Journal of Economics, 120(4) Available at SSRN: http://ssrn.com/abstract=901943.

Knack, Stephen and Philip Keefer, 1995. "Institutions And Economic Performance: Cross-Country Tests Using Alternative Institutional Measures," Economics and Politics, vol. 7(3), pages 207-227

Mauro, Paolo, 1995. "Corruption and Growth," The Quarterly Journal of Economics, MIT Press, vol. 110(3), pages 681-712, August.

McChrystal, Stanley, "COMISAF's Initial Assessment to the Secretary of Defense," August, 2009

McMillan, John and Pablo Zoido, \How to Subvert Democracy: Montesinos in Peru," Journal of Economic Perspectives, Fall 2004, 18 (4), 69 - 92

Mexico, Federal Executive Power. 1996. *Program for the Modernization of Public Administration 1995-2000,* Mexico City

http://zedillo.presidencia.gob.mx/welcome/PAGES/library/od_publicadmon.html

The News, "Pakistan sees remarkable growth in SMS traffic," April 1, 2011

The News , "Punjab Model of Proactive Governance Launched" May 30, 2010

The News, "Model citizens" April 15, 2010

Olken, Benjamin (2007) "Monitoring Corruption: Evidence from a Field Experiment in Indonesia," The Journal of Political Economy *115 (2):* 200-249

Official Communication. "Subject: Proactive Governance in Punjab (Citizen Feedback Model)," February 17, 2011

Official Communication. "Systems-Oriented Health Investment Program," September 4, 2008

Pande, Rohini, “Can Mandated Political Representation Increase Policy In uence for Disadvantaged Minorities? Theory and Evidence from India," American Economic Review, 2003, September

Rose-Ackerman, Susan, “The economics of corruption," Journal of Public Economics, 1975, 4 (2), 187-203

Svensson, Jakob. 2005. "Eight Questions about Corruption," Journal of Economic Perspectives, 19(3): 19–42.

Tella, Rafael Di and Ernesto Schargrodsky, “The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires,” Journal of Law and Economics, 2003, 46 (1), 269{292.

World Bank, World Development Report 2011: Conflict, Security, and Development April 2011.