UNIVERSITY OF CALIFORNIA, SAN DIEGO

Improving Governance and Wellbeing in Fragile States: Three Essays

A dissertation submitted in partial satisfaction of the

requirements for the degree

Doctor of Philosophy

in

Economics

by

Arman Rezaee

Committee in charge:

Professor Eli Berman, Chair
Professor Michael Callen
Professor Julie Cullen
Professor Clark Gibson
Professor Edward Miguel
Professor Karthik Muralidharan

2016

The dissertation of Arman Rezaee is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2016

DEDICATION

To Janna.

EPIGRAPH

The pen is mightier than the sword if the sword is very short, and the pen is very sharp.

—Terry Pratchett

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Top thanks to Eli Berman, followed by Michael Callen. The list goes on from there—Karthik Muralidharan, Ted Miguel, Julie Cullen, Clark Gibson, Craig McIntosh, Prashant Bharadwaj, and so many more supportive faculty at UCSD and UC Berkeley.

Thanks to all of the RAs who made this work possible—Amanullah Haneef, Umair Khawaja, Zia Mehmood, and Zarak Sohail most of all.

Thanks to everyone who provided helpful comments on these works at seminars, at conferences, and in meetings.

Thanks to my family, friends, and colleagues for their support along the way.

Chapter 1, in part, is currently being prepared for submission for publication of the material. Rezaee, Arman; Khan, Yasir; Hasanain, Ali. This research was supported by the University of California Office of the President UC Lab Fees Research Program Grant ID No. 23855, by funding from the Abdul Latif Jameel Poverty Action Lab and the Center for Effective Global Action through the Agricultural Technology Adoption Initiative, and by the International Growth Centre. Support for Rezaee's time was provided by AFOSR # FA9550-09-1-0314 and ONR # N00014-14-1-0843.

Chapter 2, in part, is currently being prepared for submission for publication of the material. Callen, Michael; Gulzar, Saad; Khan, Yasir; Hasanain, Ali; Rezaee, Arman. Support was provided by the International Growth Centre (IGC) state capabilities program, the IGC Pakistan country office, and the University of California Office of the President Lab Fees Research Program Grant #235855. Callen was supported by grant #FA9550-09-1-0314 from the Air Force Office of Scientific Research.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Callen, Michael; Gulzar, Saad; Rezaee, Arman; Shapiro, Jacob. Shapiro acknowledges support from AFOSR, grant #FA9550-09-1-0314.

VITA

2007	B.A. in Rhetoric *with honors*, University of California, Berkeley
2009	M.P.P., John F. Kennedy School of Government, Harvard University
2016	Ph.D. in Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Improving Governance and Wellbeing in Fragile States: Three Essays

by

Arman Rezaee

Doctor of Philosophy in Economics

University of California, San Diego, 2016

Professor Eli Berman, Chair

This dissertation looks at how governance and citizen wellbeing can be improved in fragile states. The first chapter examines, through a field experiment, how measuring and providing citizens information on government veterinarian performance in rural Punjab, Pakistan can improve subsequent service delivery. The second chapter examines how personality traits of three levels of government health workers predict status quo performance, response to an e-monitoring policy reform, and response to actionable information on subordinate performance. The third chapter examines how technological change during the Green Revolution influenced the decision of Pakistani leaders to govern large parts of the country at all.

Chapter 1

Crowdsourcing government accountability: Experimental evidence from Pakistan

1.1 Introduction

Asymmetric information between citizen principals and service-providing agents often leads to sub-optimal outcomes for the rural poor across the developing world (World Bank, 2004; Wild et al., 2012). In the case of *government agents*, asymmetric information has led to corruption in elected officials (Ferraz and Finan, 2011), waste in government processes (Bandiera et al., 2009), leakage between public service allocations and expenditures (Reinikka and Svensson, 2004), and more generally poor public service delivery across sectors, countries, and even continents (Chaudhury et al., 2006). In the case of *private agents*, asymmetric information has led to inefficient market allocations and rent capture at the expense of consumers (Jensen, 2007; Svensson and Yanagizawa, 2009; Aker, 2010).

Monitoring can decrease asymmetric information, but it is particularly costly to implement monitoring schemes in rural developing settings. This is because poor infrastructure makes information collection and transmission expensive in these contexts. In addition, research shows monitoring may not be effective without complimentary financial incentives (Duflo et al., 2012) and its effects attenuate as agents find alternative strategies to pursue rents (Olken and Pande, 2012).

Information clearinghouses, such as yelp.com, angieslist.com, and amazon.com, decrease asymmetries inexpensively. These crowdsourcing websites collect, aggregate, and disseminate masses of ratings at costs much lower than traditional reviewers such as the New York Times, though to date, their application has been limited to commercial settings. Furthermore, such sites have yet to take hold in the rural developing world, characterized by thin markets, low literacy rates, and 2G wireless networks.

We design and implement an information clearinghouse to reduce government agent shirking in a context fraught with asymmetric information: agricultural service provision in the developing world. Our clearinghouse provides citizens in rural Punjab, Pakistan with government veterinarians' success rates at artificially inseminating livestock, an objective measure of veterinarian effort. It gathers and disseminates locally relevant information from a large base of farmers automatically, in real time, using a call center.

Our clearinghouse model stands in contrast to government monitoring schemes that provide information to agents' superiors, relying on the "long route" of accountability in which citizens must influence policymakers to improve service provision (Callen et al., 2015). It approaches the problem more directly; it strengthens the "short route" of accountability by increasing citizens' direct power over government agents (World Bank, 2004).

And our clearinghouse strengthens government agent accountability in providing a service that is important for the livelihood of people across the developing world—renewing livestock through artificial insemination (AI). Livestock agriculture accounts for 12 percent of GDP in Pakistan, and is a key growth sector for the rural poor (Pakistan Economic Survey 2013-14). AI is crucial to renewing livestock. Most households only keep female cows because of the dual advantage of producing milk and calves, both of which require cows be pregnant. But government veterinarian shirking leads to AI success rates lower than what is possible given the technology, costing farmers potential income.

We evaluate this clearinghouse using a randomized controlled trial. Using data generated by the clearinghouse, we find that farmers treated with information on local government veterinarians' AI success rates have a 27 percent higher AI success rate than controls when they subsequently return for government services. In addition, treatment farmers are 33 percent more likely to return to a government veterinarian for AI rather than to seek a private provider.

Multiple mechanisms could explain this treatment effect on AI success rates, including treatment farmers selecting better veterinarians and/or veterinarians exerting more effort for treatment farmers. Several of our results suggest the latter—that government agents work harder when the ratings system is in place. First and foremost, treatment farmers are no more likely than control farmers to switch veterinarians after treatment. Thus the effect cannot be driven by farmers simply switching to the 'best vet' in terms of AI success and/or price. Second, treatment farmers pay lower prices after treatment.[1] While farmers may be able to improve AI success rates through their behavior alone, a change in prices requires a change in veterinarian behavior.[2]

Our estimated treatment effects on AI success are potentially subject to both selection and reporting biases since they use data from the clearinghouse. In this data, we only observe farmers who return for government AI after treatment and not those who switch to private providers, as these are not part of our clearinghouse. Returning farmers must then also choose to answer the phone and to report AI success to the clearinghouse. Importantly, we find analogous results using a representative in-person survey not subject to selection or reporting biases but with lower precision. We find an overall 26 percent treatment effect in this representative sample, which averages a treatment effect of 83 percent for farmers that select back into government AI after treatment and a treatment effect of 4 percent for attritors.[3]

[1] Note the estimated treatment effect on log AI price has a p-value of 0.12 in our primary specification.

[2] It is also possible that learning something about AI success rates in general causes farmers to take better care of their livestock and that this in turn increases AI success rates. However, we find that treatment farmers who subsequently switch to private providers do not have increased AI success rates. If our treatment effects were driven by changes in livestock care, we would expect to see effects regardless of which provider farmers subsequently choose.

[3] Note the estimated overall treatment effect has a p-value of 0.12 in our primary specification. The treatment effect for farmers that select back into government AI, analogous to the AI success rate result using clearinghouse data, is significant at 5 percent.

Our results fit the context—artificial insemination requires unobserved effort in at least two ways. First, veterinarians must keep semen straws properly frozen in liquid nitrogen canisters from the time when they are delivered to AI centers until right before insemination. Second, veterinarians must then precisely insert these straws during insemination. At the same time, farmers cannot infer a veterinarian's effort from outcomes alone. Even when executed properly, AI will not be successful 100 percent of the time, and success rates may vary based on animal health and nutrition.

In addition, while government veterinarians collect a salary and are protected from punishment for poor performance, they are legally allowed to charge a 'show-up' fee to farmers for their services on top of the fixed cost of AI. Therefore, in response to their low unobserved effort being revealed to farmers, government veterinarians may prefer to exert more effort and continue to collect a fee than to lose a customer. In other words, they may internalize the benefits of their marginal effort, a characteristic more common to private than public markets.

In a standard agency model with a stochastic outcome and inability to contract on this outcome, either unobserved agent effort (moral hazard) or unobserved inherent agent ability (adverse selection) a priori predicts both sub-optimal outcomes at baseline and that outcomes will improve as unobserved effort is revealed. We find both of these predictions to be true. However, because treatment farmers see increased AI success rates without switching veterinarians, our results rule out a pure adverse selection model and support one of moral hazard.

Several additional results from our representative in-person survey support a standard agency model. First, we find that farmers' baseline expectations about the average AI success rate of their own government veterinarians do not correlate with actual average AI success rates. This suggests the existence of asymmetric information ex ante. Second, treatment causes farmers' endline expectations about their veterinarian to become strongly correlated with the truth. This suggests that farmers indeed update their beliefs. Third, farmers who received more negative information relative to their expectations saw larger treatment effects. This suggests that the amount of information farmers receive determines their benefit.

More generally, the market for AI in rural Punjab is one in which informationally disadvantaged consumers pay more than the marginal cost of AI provision through two channels—prices and veterinarian effort. In this market, treatment-induced veterinarian effort implies consumer welfare gains so long as there are no compensating price increases or negative spillovers onto control farmers, which we do not find. Furthermore, this implies overall social welfare gains so long as the cost to veterinarians' increased effort is not too great.[4]

Our study differs from previous evaluations of the effect of information on markets with only a price channel, where changes in prices are pure transfers and any social welfare gains must come from increased market efficiency (Jensen, 2007; Svensson and Yanagizawa, 2009; Aker, 2010). Many other markets have multiple channels for rents and thus expect similar social welfare gains, including education (Andrabi et al., 2014), elections (Ferraz and Finan, 2011), and markets for private restaurants (Jin and Leslie, 2003).

In such related studies, with the exception of previous clearinghouses evaluated in Fafchamps and Minten (2012) and Mitra et al. (2014) (in both cases, the authors find no treatment effects), interventions to reduce asymmetric information are costly, static, and/or do not lead to clear social welfare gains. Our clearinghouse, on the other hand, relies on crowdsourcing technology that is cost-effective, self-sustaining, and

[4]We do not believe the marginal cost to veterinarians' increased effort induced by treatment to be very large in this setting, as travel costs are paid either way. Government veterinarians also do not spend any more time visiting treatment farmers. Any costs must be in terms of concentration, etc.

scalable. Conservative estimates suggest a 27 percent higher AI success rate translates into nearly an additional half of one month's median income per AI provided, a 300 percent return on the cost of the intervention. These effects hold out hope for improved government accountability as cellular technology improves and becomes cheaper.

The paper proceeds as follows: Section 3.2 provides background on our study district and government AI service provision there, Section 1.3 outlines our research design, including providing more information on the clearinghouse and the randomized controlled trial embedded within it, Section 3.5 provides results, Section 1.5 discusses the interpretation and social welfare implications of these results, and Section 3.6 concludes.

1.2 Background

1.2.1 The Market for AI in Sahiwal, Punjab, Pakistan

We implemented our clearinghouse in the Sahiwal district of Punjab province, Pakistan. While we selected Sahiwal based on several logistical constraints, we view it as representative of the whole of Punjab, and of similar agricultural districts across the country, though with a slightly higher prevalence of livestock.[5]

Sahiwal has a vibrant market for artificial insemination for at least two reasons. First, almost all livestock in the district are female. Second, artificial insemination decreases the costs of selectively breeding to increase milk yields, as only the semen from high-yielding bulls needs to be transported and not the bulls themselves.[6]

[5] According to the 2010 Punjab's Multiple Indicator Cluster Survey, households in Sahiwal on average have 1.4 fewer acres of agricultural land and .24 more cattle than households in other districts in Punjab. Sahiwal's average wealth, labor force participation rates, and child mortality rates are representative of Punjab.

[6] The provincial government selectively breeds livestock in two main centers in Punjab.

The government is the largest supplier in this market, offering low-cost AI services by veterinarians who have required AI training. The official cost of government AI is 50 PKR per insemination (approximately 0.5 USD), but government veterinarians are legally allowed to charge a 'show-up' fee to cover the cost of their gasoline, as well as any other costs or risks. This results in average costs of approximately 200 PKR per visit. The government has 92 one-room artificial insemination centers or veterinary offices spread throughout the district, staffed by roughly 70 active veterinarians.[7] These veterinarians' sole job is to provide artificial insemination.[8]

The only other organized supplier in this market is Nestle, but they have far fewer active veterinarians providing AI services in Sahiwal. Most private veterinarians are self-employed, buying semen from large private suppliers and providing AI services without any training. At baseline, these private veterinarians collectively provide approximately 57 percent of AI services across Sahiwal, with government veterinarians making up the remainder.

1.2.2 Asymmetric Information in the Market for AI

On a single visit, a farmer can never fully observe veterinarian effort. However, even before our intervention, farmers could have decreased asymmetries by aggregating information about their veterinarians' success rates across visits and across households. Our data suggests that they do not. At baseline, farmers' estimates of their current government veterinarian's AI success rate are uncorrelated with the truth. This can be seen in Figure 1.6, Panel A.

It then distributes the semen produced to government veterinarians across the province, including in Sahiwal.

[7]Throughout our study period, a total of 77 veterinarians were active in Sahiwal for any amount of time. Only a handful of veterinarians transferred in or out of Sahiwal.

[8]In some cases they may provide vaccinations during AI service provision, but this occurs very rarely. A smaller, distinct group of veterinarians care for sick animals.

This asymmetric information contributes to AI success rates that are lower than what veterinarians can achieve. At baseline, AI success rates average approximately 70 percent, while success rates of 85-90 percent are possible with the training and equipment in Sahiwal.

1.3 Research Design

1.3.1 The Clearinghouse

To measure veterinarian prices and effort and to subsequently disseminate that information to consumers, we developed a novel cellular-based information clearinghouse. Figure 1.1 diagrams the four components of our intervention.

Pre-treatment: During the study, government veterinarians in Sahiwal were required to collect real time information on all AI service provisions using an Android smartphone equipped with an Open Data Kit-based application.[9] The data was immediately sent to the clearinghouse. We denote this data collection as t = 0 in Figure 1.1.

Data collection and aggregation: Each service provision generated two subsequent phone calls. First, one day later (denoted t = +1 day in Figure 1.1), a representative from the clearinghouse call center called the farmer to verify that the veterinarian had provided service and to ask what price he had charged. Then, sixty days later (t = +60 days), they called again to ask if the artificially inseminated livestock were pregnant. The clearinghouse continuously aggregated this price and AI success rate data for each veterinarian.

Treatment: The clearinghouse collected and aggregated information from January to September, 2014. Treatment began in October 2014, once we had sufficient

[9] In practice, veterinarians did not always comply. See Section 1.4.3 for discussion.

data on veterinarians to have meaningful measures of price and AI success rates. Treatment took place during the second call (at t = +60). Only this time a randomized group of farmers was provided information on local veterinarians' prices and AI success rates. The uninformed farmers became the control group.

Post-treatment: The clearinghouse allowed us to link farmers over time, so we observe post-treatment government AI provision for both treatment and control farmers (if they return; Figure 1.1 depicts the return of a treatment farmer but not a control farmer). These post-treatment observations also generate two follow-up phone calls.[10]

1.3.2 Information Provision

In the treatment group, the clearinghouse representative presented farmers with information on the top three veterinarians within three kilometers of their household in terms of AI success rates for cows, and the top three veterinarians in terms of AI success rates for buffalo.[11]

We gave treatment farmers AI success rates for these three to six veterinarians, and the average price of the service, during the second follow-up call.[12] The clearinghouse then sent a follow-up SMS with the same information. If farmers requested it, we also gave them veterinarians' phone numbers, information on average farmer-reported satisfaction with veterinarians on a 1-5 scale, and information on any other veterinarian in our system.

The clearinghouse administered treatment at the farmer level through a coin-

[10]Note, however, that treatment selection is carried forward in time. See Section 1.3.2.

[11]When we had fewer than 25 observations for a veterinarian, we weighted success by $\sqrt{n}/5$, where n was the number of observations. By design, almost every veterinarian had more than 25 observations each for cows and buffalo once the treatment began. The exceptions were two veterinarians hired after our treatment began in October 2014.

[12]There can be overlap in the most successful veterinarians in terms of cows and buffalo.

flip stratified on the nearest government veterinary clinic to a farmer's household. Farmers who returned for service provision after treatment assignment retained their initial assignment. Note that treatment occurred at a different time for each farmer, 60 days after they first entered our clearinghouse. This means that the post-treatment period differs for each farmer.[13]

1.3.3 Representative Survey

In addition to the clearinghouse data, we independently surveyed a representative sample of farmers from across Sahiwal. We did so because the clearinghouse sample is not representative: to enter the clearinghouse, farmers first selected government AI over private, then their government veterinarian complied to record their service provision, then we were able to reach them on the phone to collect price and AI success information; and then we only observed post-treatment outcomes for clearinghouse farmers who subsequently returned to a government veterinarian for AI (as opposed to a private provider).

For these surveys, we sampled 90 of Sahiwal's approximately 500 villages from a district village census.[14] Within each village, we selected ten households using the Expanded Program on Immunization (EPI) cluster sampling method (Henderson and Sundaresan, 1982). We selected households that reported owning at least two livestock (cows and/or buffalo) and having regular access to a cellular phone.

We manually entered survey farmers' phone numbers into our clearinghouse to

[13]Unfortunately, the coin used for randomization was shaved, due to a glitch in the clearinghouse algorithm. This resulted in 52 percent of farmers being treated. However, the probability of treatment remained fixed across farmers across time.

[14]We stratified the sample by whether or not a government veterinarian center was in each village and on whether each village bordered an irrigation canal. The sample is representative of Sahiwal in terms of: area, settled area, cultivated area, area of wheat, rice, cotton, sugar cane, pulses, orchards, and vegetables, having a river, distance to the nearest veterinarian center, number of livestock in the village, literacy rates, religion, age, and standard wealth index characteristics. Results available upon request.

generate treatment or control follow-up calls. These calls were near identical to those to farmers that entered our clearinghouse on their own, and the treatment information provision component was identical.[15]

Sample villages can be seen in Figure 1.7. Figure 1.2 presents a timeline of the clearinghouse and survey data collection. The baseline survey occurred prior to our clearinghouse implementation, and the endline survey occurred immediately prior to the clearinghouse being shut down.[16]

Tables 1.1, 1.2, and 1.10, report the balance of our clearinghouse and representative survey samples between treatment and control farmers.

1.3.4 Empirical Specifications

We use the following specification for our primary analysis:

$$outcome_{ft} = \alpha + \beta T_f + \Gamma_{ft} + \epsilon_{ft} \tag{1.1}$$

where $outcome_{ft}$ is an outcome for farmer f from post-treatment AI visit t. T_f is a treatment indicator, Γ_{ft} are treatment strata and other baseline controls to improve precision, and ϵ_{ft} is an idiosyncratic error term. While we administered treatment at the farmer level, treatment information provision was localized at the village-cluster level. We cluster standard errors at this village-cluster level to allow for correlation in outcomes between farmers in the same village-cluster. Village-clusters are groups of villages that share the same government veterinarians within a three kilometer

[15]The only difference was that instead of asking questions about a specific recorded service provision from 60 days ago as is the case with clearinghouse calls, we asked about farmers' last AI service.

[16]We conducted a purely technical survey at midline to collect new phone numbers for those households that changed numbers between the baseline and the first round of treatment phone calls. This allowed us to treat as many independently surveyed farmers as possible.

radius. There are roughly two villages per village-cluster.

We define post-treatment for control farmers as all observations after the phone call in which they were selected into control rather than treatment. This ensures balance in the length of the post period between treatment and control farmers.

We have four primary outcomes:

$Switched\ veterinarians_{ft}$: a dummy variable equal to one if a farmer's veterinarian at visit t differed from the farmer's veterinarian at visit $t-1$.

$Log\ price_{ft}$: the log price paid for AI at visit t, as reported by the farmer when called the next day.

$AI\ success\ rate_{ft}$: a dummy for the success of the AI provided at visit t, as reported by the farmer when called 60 days later.

$Returned_f$: a dummy variable equal to one if a farmer returned for government AI after treatment by the end of the project.[17]

1.4 Results

In this section, we present results. First, we present treatment effects using our representative sample (Section 1.4.1) and our clearinghouse sample (Section 1.4.2). Second, we show that treatment does not induce veterinarian reporting bias (Section 1.4.3) or farmer reporting or selection biases (Section 1.4.4) in the clearinghouse sample. Third, we explore the primary mechanism for our treatment effects, decreased moral hazard or increased effort by veterinarians for the treated, through heterogeneity analyses (Sections 1.4.5 and 1.4.6).

[17]We pre-specified our empirical specification in our pre-analysis plan, registered in the AEA RCT registry. We did not pre-specify $Returned_f$. We did pre-specify $Switched\ veterinarians_{ft}$, $Log\ price_{ft}$, and $AI\ success\ rate_{ft}$. We pre-specified the latter two outcomes conditional on veterinarian switching, but we have made them unconditional since we do not observe veterinarian switching.

1.4.1 Treatment Effects—Representative Sample

Table 1.3 presents treatment effects using our representative sample. We report first effects on price. Column (3) shows a statistically insignificant price reduction for the entire sample, which remains insignificant if we disaggregate into the subsamples of farmers who either returned to government AI (1) or attritted to private providers (2) after treatment. In column (4), we find that treatment farmers who return to government AI have a 47 percentage point, or 83 percent, higher AI success rate. In contrast, column (5) reports an insignificant treatment effect on AI success for farmers who attrited, indicating that treatment does not induce farmers to seek out a better private provider. In column (6) we find that, while it is not quite significant, overall AI success rates are large and positive even when including those farmers that attrited: treatment farmers have a 17 percentage point, or 26 percent, higher AI success rates after treatment.[18]

While these results are not subject to reporting or selection biases, the size of our representative sample allows for less precision than with our clearinghouse sample, which we will now turn to.

1.4.2 Treatment Effects—Clearinghouse Sample

Table 1.4 presents treatment effects of information provision on our primary outcomes using the clearinghouse sample. In column (1), treatment farmers are 3.2 percentage points, or 33 percent, more likely than control farmers to return for government AI after treatment.[19] As a visualization, we present an added-variable plot

[18]The p-value of this estimate is 0.12.

[19]The low overall return rate is likely because the average time for farmers between treatment and the end of our study period is five months and AI is only required roughly once a year per animal. As we see in Table 1.5 as well, only 30 percent of return visits were recorded by veterinarians, so even in five months the true return rate is likely 40 to 50 percent.

of this result in Figure 1.3.

In columns (2) through (4), we present effects on those farmers that return after treatment selection. In columns (2) and (3) we find that there are no statistically significant treatment effects on veterinarian switching or on log prices, though the coefficient on log price is nearly significant with a p-value of 0.12. In column (4), we find that treatment farmers have a 17 percentage point, or 27 percent, higher AI success rate after treatment.

This treatment effect on AI success rates is substantially smaller in magnitude than the analagous 47 percentage point treatment effect we report in Table 1.3, column (4) in the representative sample. However, we cannot reject that the effect in the representative sample is equal to that in the clearinghouse sample.

In Figure 1.4, we present the treatment effect on AI success rates in real time (as opposed to in pre/post time, where post begins at a different time for each farmer). The top panel illustrates that treatment farmers have higher AI success rates consistently over time, while the bottom panel traces the size and significance of this treatment effect over the post period. These results suggest that any information spillovers between treatment and control farmers are either small or fixed throughout time. The latter is unlikely given the rolling nature of treatment. If anything, there is a small bump up in AI success rates for control farmers in the first month of the treatment, which suggests positive information spillovers. This would attenuate our results. The figure also suggests that there are no negative spillovers onto control farmers from veterinarian effort constraints.

The most likely cause of the across-the-board downward trend in AI success rates beginning in March 2014 is changes in leadership of the Punjab Livestock and Dairy Development Department at both the provincial and Sahiwal district levels—the new regime was less focused on veterinarian performance than the last had been.

In Figure 1.5, we present the treatment effect on log AI prices in real time. We find that the same visual trends hold for prices, and that when we bootstrap standard errors, the treatment effect is significant in six of eight months.

We reproduce our primary treatment effects on our representative survey sample, selecting on returning for government AI after treatment, in Table 1.11. The point estimates are of a similar magnitude.[20]

1.4.3 Treatment Does Not Induce a Veterinarian Reporting Bias

In order to believe the internal validity of our clearinghouse sample, it is important to note in Table 1.5 that treatment does not induce a reporting bias among government veterinarians. We measure reporting bias by comparing farmer reports of service provision from our representative survey with entries in the clearinghouse. While government veterinarians only comply by reporting AI approximately 30 percent of the time, they are equally likely to report for treatment and control farmers.

1.4.4 Ruling Out Farmer Selection and Reporting Biases in the Clearinghouse Sample

For the same reason, we must also rule out farmer selection and reporting biases. Our estimates would include a farmer selection bias if farmers that would otherwise see higher success rates are those that select back into government AI after treatment. Our estimates would include a farmer reporting bias if treatment farmers are more or less likely to answer the phone when we call to ask about AI success.

[20]Note that the mean return rate of control farmers is higher in this sample, but not three times that of the clearinghouse sample. This is consistent with the fact that we do not rely on veterinarian reporting for this data. Also, these farmers had less time after treatment to return to our sample on average.

We have already presented evidence against both farmer selection and response biases in Table 1.3, column (6). Accounting for attriters removes possible selection bias. In addition, the representative survey had a successful follow-up rate of 96 percent with no differential attrition, which removes possible response bias.

As an additional check for farmer selection bias, in Table 1.6 we show balance on all measured pre-treatment outcomes, including AI success rates, between returning treatment and control farmers in the clearinghouse data. While this does not rule out selection on unobservables, we believe that it does rule out the most likely type of selection that could drive such a large increase in AI success rates in our post-treatment sample—selection back into government AI by farmers who have younger, healthier livestock more likely to get pregnant. If this selection were occurring, such younger and healthier animals should have then been more likely to get pregnant in the pre-periods as well, yet we do not see this. We also do not see any differences in past prices paid, past veterinarian switching, or other administrative variables.

1.4.5 Treatment Effects by Government Veterinarian Rank

In order to explore the mechanism for our treatment effects, we present a series of heterogeneous treatment results that support a standard moral hazard model.

First, in Table 1.7, we present treatment effects for two important sub-populations, separated according to the ranking of the last government veterinarian who served them—those for whom this veterinarian was ranked in the top three in their village-cluster, and those for whom he was not. This aligns with those veterinarians on whom treatment farmers received information regarding AI success rate and price. We separate control farmers based on what they would have been told, had they been treated.[21]

[21]Note that at the beginning of our treatment phone calls we verify farmers' villages as they were automatically generated by GPS. This verification is not done with control

We find suggestive evidence that our main results are localized to farmers whose past veterinarian was not ranked in the top three in their area at the time of treatment.[22] Again, this is in line with a standard moral hazard model. The more a farmer learns a veterinarian can increase unobserved effort, the more s/he is able to then bargain away rents from the veterinarian.[23]

Perhaps the most surprising result in Table 1.7 is that farmers whose past veterinarian was not ranked in the top three are more likely to return. To investigate this, we show in Table 1.12 that farmers in Table 1.7 Panel B tend to live almost twice as far away from their closest veterinary center.[24] This is consistent with farmers living in more remote areas settling for lower effort veterinarians because of higher switching costs. And it is exactly these farmers with higher switching costs that receive the largest benefits from treatment.

1.4.6 Results Using Farmer Expectations from the Representative Survey Sample

If we are to believe that our results are in line with a standard moral hazard model, we should expect the level of asymmetric information between farmers and veterinarians at baseline to be important. We present three results in this vain, in this case using farmers' stated expectations. These expectations come from our

farmers. To avoid measurement error correlated with treatment, we separate treatment farmers based on what they would have been told had we not verified their village. This hypothetical information set correlates with the truth at over 90 percent.

[22]These results are suggestive because, while the point estimates are qualitatively different, we cannot reject this difference with significance.

[23]We should also expect heterogeneous treatment effects based on whether or not a farmer's past government veterinarian was ranked top in their village-cluster versus second best, or second best versus third best, etc. We do not have power to accurately detect these differences, but results are consistent with the same simple model. Results available upon request.

[24]In addition, these farmers have more buffalo. We control for baseline means of both of these variables in Table 1.7.

representative survey sample, in which we asked farmers what they expect the average AI success rate of their past veterinarians to be.

In Figure 1.6, we compare farmers' expected average AI success rate for their veterinarian prior to treatment with the actual average AI success rate of that veterinarian. Actual average AI success rates are drawn from our clearinghouse data prior to October 2014 when treatment calls began.

Our first result is in Panel A of the figure—at baseline there is no correlation between farmer expectations and the true AI success rate of their veterinarian. This suggests there is room to improve service delivery by relieving asymmetric information.

Our second result is in Panel B of the figure—at endline there is a strong correlation between expectations and the truth for treatment farmers. In other words, treatment changes expectations. This is a crucial test that information was passed on through our treatment. Panel C presents the endline correlation for control farmers—while much smaller than with treatment farmers and insignificant, there is a positive correlation. Thus suggests potential information spillovers between treatment and control farmers, which would attenuate our treatment results above.

Point estimates for these two results are reported in Table 1.8. The null hypothesis that the coefficients in columns (2) and (3) are equal is almost rejected, with a p-value of 0.115.

Third, using farmer expectations we can also separate treatment effects by the level of asymmetric information between farmers and veterinarians at baseline. To do so, we difference farmers' expected average AI success rate with the truth. We then split our sample according whether farmers had above or below the median in this difference. Positive values in this difference occur when farmers are told that their veterinarian is better than they expected; negative values occur when farmers are told their veterinarian is worse than they expected. The median is .012.

Table 1.9 presents results from this heterogeneity analysis. We find that, as with treatment effects by government veterinarian rank, the more unexpectedly negative the information a farmer receives about their veterinarian, the more s/he is able to then bargain away rents from the veterinarian.

1.5 Discussion

1.5.1 Interpretation: Unobserved Effort or Inherent Ability?

Several results suggest that the treatment effect on AI success rates is entirely due to increased veterinarian effort for the treated. To illustrate this, we can walk through the process by which farmers select a veterinarian and negotiate prices and effort. First, farmers decide whether to get AI at all when a cow is in heat. Next, they decide whether to stick with their previous veterinarian. If farmers switch, they then decide whether to call a government or private veterinarian. Finally, they decide how to engage with this veterinarian in pre-visit negotiations over the phone as well as during the AI visit (and veterinarians have to decide how to respond).

In our setting, farmers almost always choose to inseminate their livestock in heat, so we would not expect any changes in this decision. Next, we show in Table 1.4 that treatment farmers are no more likely than control farmers to switch veterinarians after treatment. Thus the treatment effect cannot be driven by farmers simply switching to the 'best vet'.

We do see changes in whether farmers call a government or private veterinarian, however. Importantly, we show in Table 1.3 that treatment farmers who subsequently switch to private providers do not have increased AI success rates. If our treatment effect is driven by changes in farmer behavior towards their livestock, we would expect effects regardless of which veterinarian the farmer selects after treatment. The same argument can be applied to the results from Section 1.4.5. If our treatment effect is driven by changes in farmer behavior, farmers' past veterinarian ranking should not matter.

Thus, we can turn to the final part of the decision process as the likely mechanism—farmers' engagement with veterinarians. Our results are consistent with farmers using the information we provide to them to negotiate reductions in government veterinarians' informational rents through higher effort and lower prices. And while farmers may be able to improve AI success rates through their behavior alone, the decrease in prices that we find requires a change in veterinarian behavior.

If we are to view increased veterinarian effort as the driver of our results, then that effort must be easily varied across visits. Anecdotes suggest that this is true. One commonly cited example of low veterinarian effort is the way in which veterinarians treat semen straws. As mentioned above, the provincial government delivers these straws to veterinary centers in liquid nitrogen canisters, and they must be kept frozen until just before use. Veterinarians sometimes take straws out before leaving on a visit rather than transporting the canister to the farm. This likely results in the semen spoiling, though the veterinarian still performs AI and charges the farmer. And because farmers call veterinarians before AI to negotiate a time and price, treatment farmers could pressure them to take better care transporting semen. Veterinarians would have to exert more effort but farmers would likely still pay them positive rents rather than having to pay the cost to find a new veterinarian.

1.5.2 Social Welfare Implications

To understand the social welfare implications of this intervention, we consider benefits and costs to farmers and to veterinarians as well as the cost of the intervention itself.[25]

Benefit to farmers: if the treatment effect of 27 percent on AI success rates translates into just three percent more calves born per year per farmer (i.e., if farmers with a failed AI attempt are able to successfully impregnate their animal two months later), and the expected value of a calf is roughly 107,500 PKR (approximately 1075 USD) at the market, then treatment farmers would earn an additional 3,225 PKR (32 USD) per year, equal to nearly half of one month's median income.[26] This is a conservative estimate. It does not count the additional net value of two months of milk nor the cumulative net present value effect of an increased future stream of livestock.

Cost to farmers: we showed that farmer treatment effects are not due to changes in farmer behavior, we do not consider there to be costs to farmers of this intervention.

Benefit to veterinarians: farmers do not switch veterinarians more as a result of treatment, which suggests no change in veterinarian market shares that could impact social welfare. However, treatment farmers are more likely to return for government AI. Thus, if anything, government veterinarians benefit from this intervention. This would be at the cost of private veterinarians, however, so we will not consider it.

[25] We do not consider changes in price as such is a transfer with no net social welfare implications.

[26] This calf value is the average of male and female calf prices reported at http://www.pakdairyinfo.com/feasibility.htm, accessed 10/8/2015. The monthly median income of households in Paksitan, according to the World Bank, is 73.26 USD per month, accessed 10/8/2015.

Cost to veterinarians: we do not believe the marginal cost to veterinarians' increased effort induced by treatment to be very large in this setting, as travel costs are paid either way. Government veterinarians also do not spend any more time visiting treatment farmers. Any costs must be in terms of concentration, etc.

Cost of the intervention: including one-time fixed costs to develop our clearinghouse technology, this intervention cost approximately 50,000 USD to reach over 6,000 farmers for treatment or control calls, or approximately 8 USD per farmer.

Adding it up, we find benefits of 32 USD per farmer from an intervention that cost 8 USD per farmer. This suggests a large, 300 percent return.

1.6 Conclusion

In this paper, we present results from the randomized controlled trial of a novel solution to a common government accountability failure: shirking by government agents in a setting of asymmetric information. Our solution is novel not only in that it leverages the cost-effective, self-sustaining nature of crowdsourcing to help the poorest, but also in that it does so in a tough setting. In rural Punjab, the market for artificial insemination is thin, literacy rates are low, and cellular networks are very limited—yet we were able to employ an information clearinghouse with success.

The very fact that our clearinghouse was successful purely through providing information confirms the existence of asymmetric information in this setting. And the fact that veterinarians respond with increased effort confirms that this asymmetric information is about unobserved effort. While these confirmations are neither novel nor heartening in and of themselves, they allow us to fit the livestock sector in Punjab into a context that is much more general. Moral hazard has been documented in numerous sectors, public and private, across the developing world. We might expect our clearinghouse to help citizens in any of these sectors, so long as they answer the phone.

And given the low cost of our clearinghouse, we might expect similarly large returns in other sectors. Conservative estimates suggest a 300 percent return to farmers on the cost of the intervention. This is driven by a 27 percent increase in AI success rates for treatment farmers. In other words, thousands of poor, rural Pakistanis who were treated are now more likely to have milk to drink and calves to raise or to sell for substantial income. This is heartening.

As a testament to the scalability of our clearinghouse, we have already begun conversations within the Livestock Department about expanding the program to all of Punjab. This would require no additional fixed costs and less than proportional marginal costs. Across contexts, we are already experimenting with scaling our information clearinghouse to relieve asymmetric information between citizens and pollution regulators in Punjab. We hope to learn how crowdsourcing can work in a regulatory rather than a market environment, and for public rather than private goods.

Some of these effects of improving the flow of information depend on how connected the population is, and the price of connectivity. In another project, we are experimenting with the placing of cellular towers to understand how economic, social, and political outcomes are impacted by across-the-board decreases in the cost of transmitting information.

We hope this paper and other new studies will improve our understanding of how technology can be leveraged to improve the feasibility and impact of already tried-and-true interventions, such as monitoring to reduce asymmetric information. As cellular networks improve and as technology to collect, aggregate, and disseminate information advances, our results suggest we may see improved outcomes for citizens across the rural developing world.

1.7 Acknowledgements

Chapter 1, in part, is currently being prepared for submission for publication of the material. Rezaee, Arman; Khan, Yasir; Hasanain, Ali. This research was supported by the University of California Office of the President UC Lab Fees Research Program Grant ID No. 23855, by funding from the Abdul Latif Jameel Poverty Action Lab and the Center for Effective Global Action through the Agricultural Technology Adoption Initiative, and by the International Growth Centre. Support for Rezaee's time was provided by AFOSR # FA9550-09-1-0314 and ONR # N00014-14-1-0843.

1.8 Chapter 1 Appendix

1.8.1 Tables and Figures

Table 1.1: Treatment Balance—Clearinghouse Data

	Treatment	Control	Difference	P-value
Satisfaction with AI service provision (1-5)	4.185	4.136	0.049	0.123
	[0.736]	[0.760]	(0.029)	
Farmer switched vets since last AI visit	0.052	0.047	0.005	0.133
	[0.222]	[0.213]	(0.0100	
AI visit charges (PKR)	196	203	-7	0.479
	[180]	[250]	(9)	
AI visit success rate (pregnancy / AI attempts)	0.686	0.687	-0.002	0.432
	[0.458]	[0.457]	(0.016)	
No of cows owned by farmer	2.544	2.447	0.097	0.312
	[3.439]	[3.053]	(0.155)	
No of buffalo owned by farmer	3.121	3.315	-0.195	0.771
	[3.777]	[6.347]	(0.366)	
Distance to closest AI center (km)	2.170	2.277	-0.107	0.825
	[2.254]	[2.259]	(0.114)	

Notes: Standard deviations reported in brackets. Standard errors reported in parentheses. Means and differences are unconditional. P-values are from OLS regressions with randomization strata fixed effects and standard errors clustered at the village-cluster level. The sample consists of 6,473 pre-treatment farmer-visit-level observations from 3,094 unique farmers across 202 village-clusters. Some regressions have fewer observations due to missing data. Beginning in October 2014, treatment farmers received information about the AI success rates of their local government veterinarians. Satisfaction, AI visit charges, and numbers of cows and buffalo are reported by farmers on the phone one day after AI service provision. AI visit success rate is reported by farmers on the phone 60 days after AI service provision. Farmer switched vets and distance to closest AI center are automatically generated administrative data.

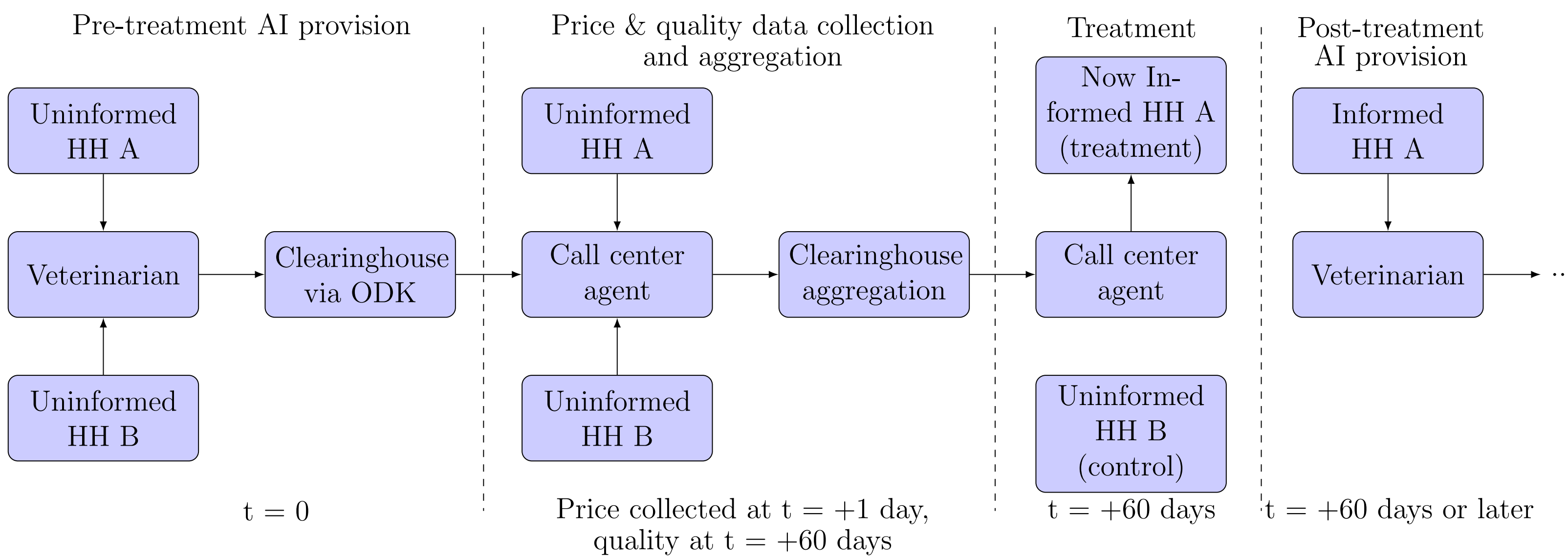

Figure 1.1: Clearinghouse Flowchart

Notes: Arrows indicate the flow of information. The collection of quality data and treatment occur during the same follow-up phonecall 60 days after service provision. Beginning in October 2014, treatment farmers received information about the AI success rates of their local government veterinarians.

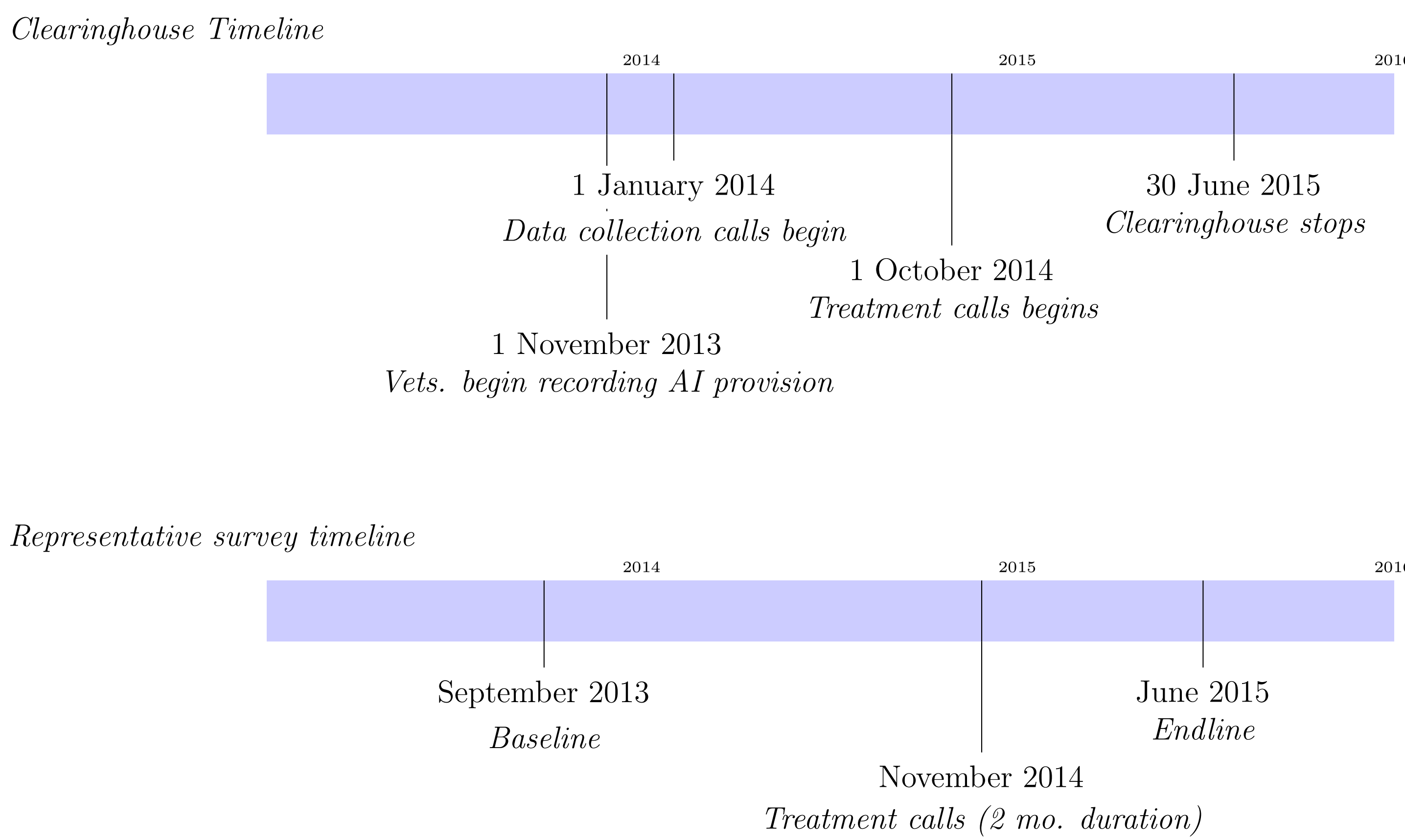

Figure 1.2: Clearinghouse and representative survey timelines

Table 1.2: Treatment Balance—Representative Survey Sample

	Treatment	Control	Difference	P-value
Farmer-level baseline variables—190 observations across 61 village-clusters				
Livestock is primary source of HH's income (=1)	0.085	0.097	-0.012	0.748
	[0.281]	[0.297]	(0.042)	
1-10 effort household puts into selecting veterinarian	6.200	5.575	0.625	0.491
	[2.361]	[2.049]	(0.537)	
Farmer attrited from in-person endline	0.021	0.011	0.011	0.812
	[0.145]	[0.104]	(0.018)	
Farmer-visit-level variables—356 pre-treatment observations from 190 farmers across 61 village-clusters				
Farmer switched vets since last recorded AI visit (=1)	0.179	0.190	-0.011	0.879
	[0.385]	[0.393]	(0.055)	
AI visit charges	367	356	10	0.771
	[373]	[361]	(48)	
AI visit success rate	0.703	0.750	-0.047	0.159
	[0.447]	[0.431]	(0.049)	
1-10 AI visit farmer satisfaction	7.694	9.302	-1.608	0.290
	[2.184]	[22.333]	(1.754)	
1-10 farmer estimated AI visit veterinarian success rate	6.636	6.315	0.321	0.606
	[1.739]	[1.981]	(0.276)	

Notes: Standard deviations reported in brackets. Standard errors reported in parentheses. Means and differences are unconditional. P-values are from OLS regressions with randomization strata fixed effects and standard errors clustered at the village-cluster level. Some regressions have fewer observations due to missing data. All data come from baseline surveys fielded in August and September 2013, with the exception of "Farmer attrited from endline survey". This variable is a dummy equal to one if a farmer was present during our baseline survey and not our endline survey. The sample of farmers was selected to be geographically representative of Sahiwal and is drawn from 90 different villages. The sample is limited to farmers that report receiving services from a government veterinarian at baseline. Treatment farmers received information about the AI success rates of their local government veterinarians. Treatment calls were conducted in November 2014 and January 2015.

Table 1.3: Treatment Effects—Representative Survey Sample

Outcome:	Log price			AI success rate		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment farmer (=1)	0.027	-0.146	-0.062	0.470**	0.028	0.172
	(0.405)	(0.216)	(0.164)	(0.186)	(0.187)	(0.109)
Mean of dependent variable	5.856	5.888	5.874	0.567	0.765	0.672
# Observations	69	87	156	63	79	142
# Village-clusters	27	39	53	29	35	51
R-Squared	0.633	0.655	0.540	0.498	0.281	0.271
Sample	Returned	Attrited	Both	Returned	Attrited	Both

Notes :*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors clustered at the village-cluster level reported in parentheses. All regressions include randomization strata fixed effects, survey wave fixed effects, and controls for baseline mean outcomes (switched veterinarians, log price, AI success rate) as well as dummies for whether the given observation is missing each baseline mean outcome. The sample is limited to post treatment reports of AI service provision from farmers during our endline survey, conducted in June 2015. Treatment farmers received information about the AI success rates of their local government veterinarians. Treatment calls were conducted in November 2014 and January 2015. Returned indicates farmers that received government AI before treatment and subsequently returned for government AI after treatment by the end of the project. Attrited indicates farmers who received government AI before treatment and instead subsequently received private AI by the end of the project. Log price and AI success rates are recalled by farmers from service provisions two to seven months ago.

Table 1.4: Treatment Effects—Clearinghouse Data

Outcome:	Returned	Switched veterinarians	Log price	AI success rate
	(1)	(2)	(3)	(4)
Treatment farmer (=1)	0.032***	0.007	-0.270	0.168**
	(0.011)	(0.028)	(0.170)	(0.083)
Mean of dependent variable	0.098	0.084	5.248	0.623
# Observations	3184	629	312	240
# Village-clusters	205	111	103	98
R-Squared	0.192	0.305	0.596	0.489
Sample	Pre	Post	Post	Post

Notes :$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the village-cluster level reported in parentheses. All regressions include randomization strata fixed effects and controls for baseline mean outcomes (switched veterinarians, log price, AI success rate) as well as dummies for whether the given observation is missing each baseline mean outcome. In addition, columns (2) through (4) include exact call center script fixed effects and a time trend control. The sample for column (1) is farmers that received a government AI service and were subsequently treated, regardless of whether they then returned. The sample for columns (2) through (4) are farmers that returned after treatment. Note the differences in observations across columns are due to the fact that veterinarian switching can be detected without any successful phone calls, where as log price requires one successful phone call and AI success rate requires two successful phone calls to a farmer. Beginning in October 2014 treatment farmers received information about the AI success rates of their local government veterinarians. Returned is a dummy variable equal to one if a farmer that received government AI before treatment subsequently returned for government AI after treatment by the end of the project. Switched veterinarians is a dummy variable equal to one if the veterinarian that a farmer saw for a service provision was different than the last veterinarian seen. Log price is the log price paid for the service provision, as reported by the farmer when called to verify service provision. AI success rate is the rate of success of the AI services provided at a specific service provision upon follow up 60 days later.

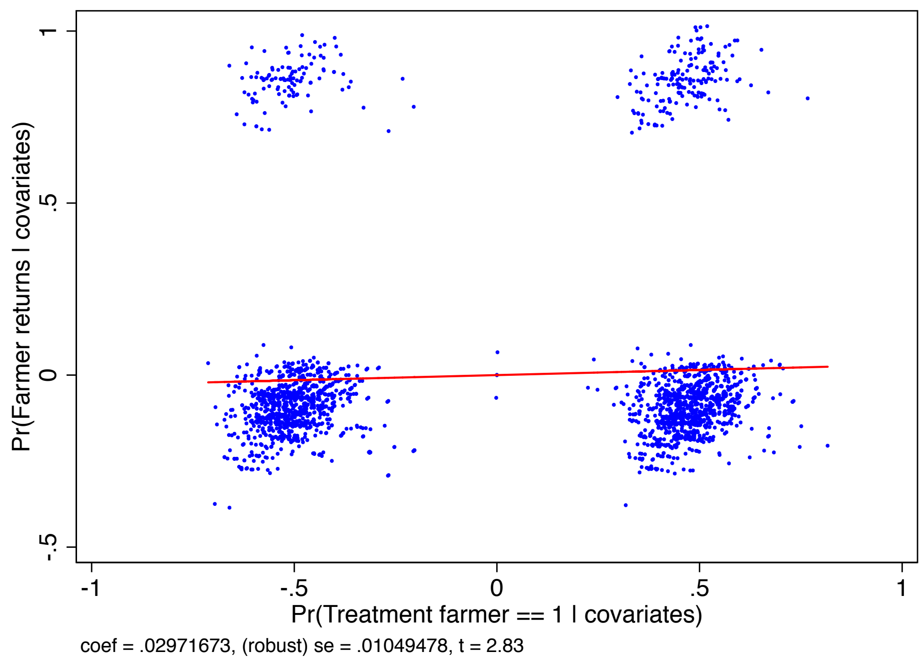

Figure 1.3: Farmer Returned Added-variable Plot—Clearinghouse Data

Notes: The sample is farmers that received a government AI service and were subsequently treated, regardless of whether they then returned. Beginning in October 2014 treatment farmers received information about the AI success rates of their local government veterinarians. Returned is a dummy variable equal to one if a farmer that received government AI before treatment subsequently returned for government AI after treatment by the end of the project. The covariates used to predict residual values are randomization strata fixed effects and controls for baseline mean outcomes (switched veterinarians, log price, AI success rate) as well as dummies for whether the given observation is missing each baseline mean outcome.

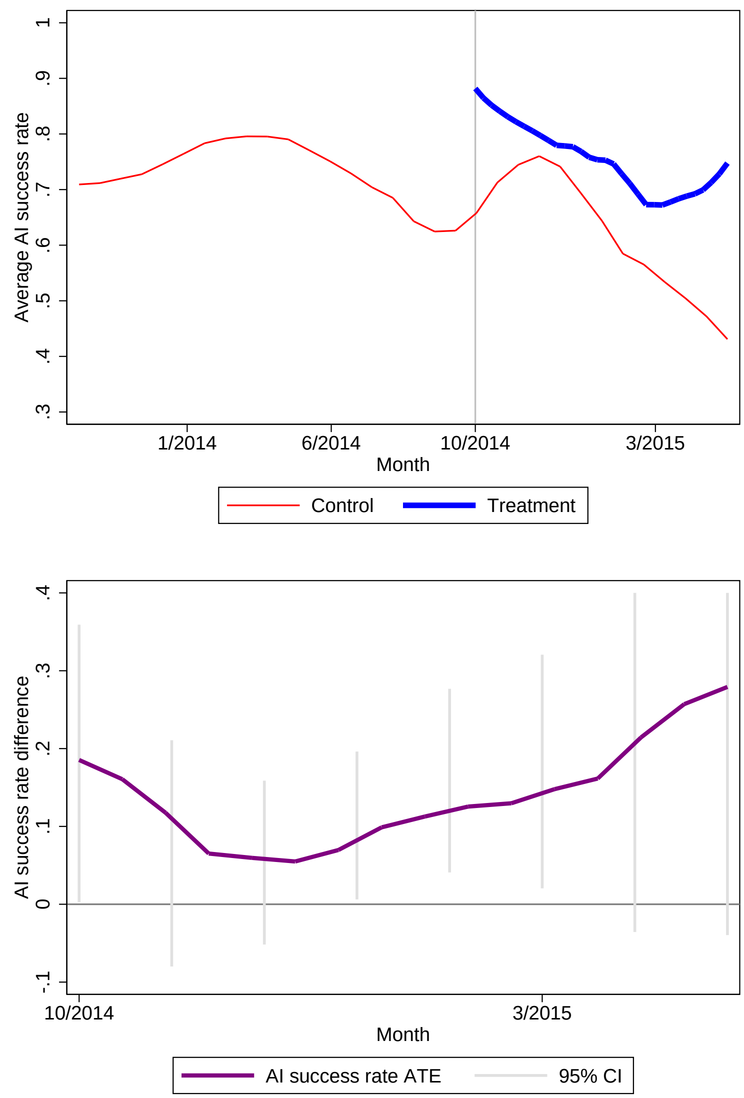

Figure 1.4: AI Success Rates in Real Time—Clearinghouse Data

Notes: The sample is farmers that received a government AI service and then answered the phone and reported AI success 60 days later. Beginning in October 2014 treatment farmers received information about the AI success rates of their local government veterinarians. Lines are smoothed using a kernel-weighted local polynomial regression with the Epanechnikov kernel and bandwidth one. Confidence interval bootstrapped and truncated at 0.4.

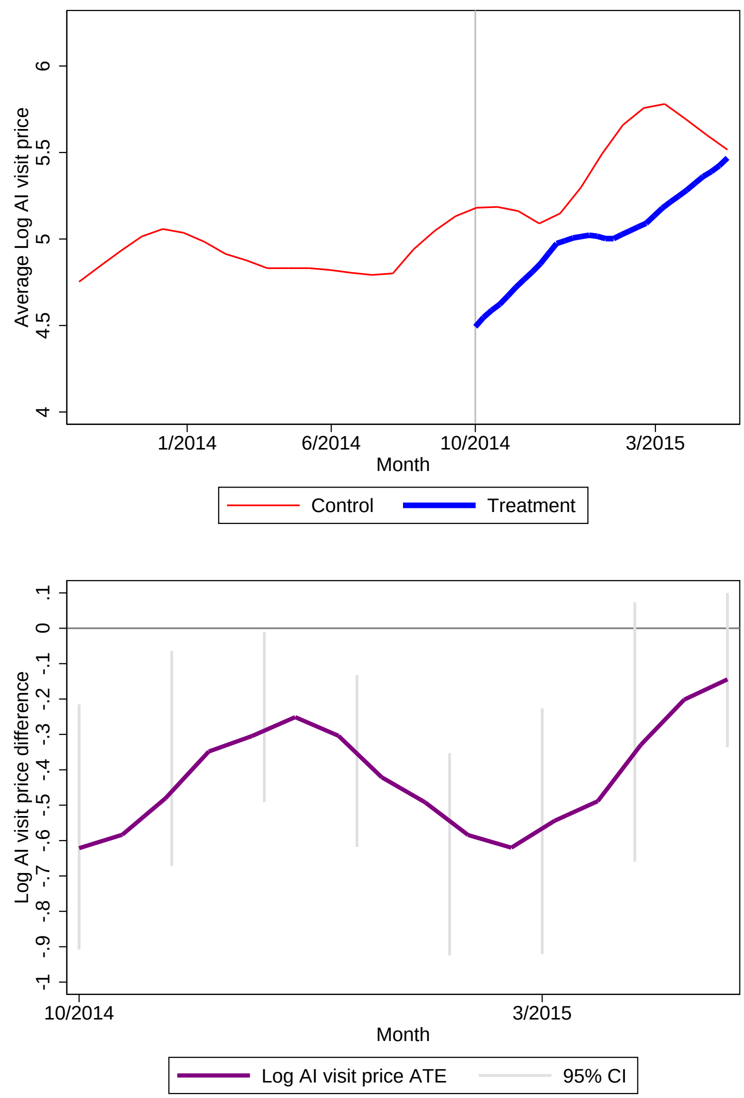

Figure 1.5: Log Price per AI Visit in Real Time—Clearinghouse Data

Notes: The sample is farmers that received a government AI service and then answered the phone and reported price paid one day later. Beginning in October 2014 treatment farmers received information about the AI success rates of their local government veterinarians. Lines are smoothed using a kernel-weighted local polynomial regression with the Epanechnikov kernel and bandwidth one. Confidence interval bootstrapped and truncated at 0.1.

Table 1.5: Does Treatment Induce a Veterinarian Reporting Bias?

	Treatment	Control	Difference	P-value
Farmer reported AI and veterinarian submitted data to call center (=1)	0.299	0.276	0.023	0.758
	[0.459]	[0.448]	(0.044)	.
Farmer reported receiving a call verifying AI service (=1)	0.287	0.240	0.047	0.566
	[0.449]	[0.422]	(0.041)	.

Notes: Standard deviations reported in brackets. Standard errors reported in parentheses. Means and differences are unconditional. P-values are from OLS regressions with randomization strata fixed effects and standard errors clustered at the village-cluster level. The sample consists of 730 farmer-visit-level observations from 440 unique farmers across 83 village-clusters from our endline survey, conducted in June 2015. Some regressions have fewer observations due to missing data. Treatment farmers received information about the AI success rates of their local government veterinarians. Treatment calls were conducted in November 2014 and January 2015."Farmer reported AI and veterinarian submitted data to call center" is a dummy equal to one if a government AI service provision reported in our endline survey was subsequently submitted to the clearinghouse by the veterinarian that performed the service. This is done by verifying survey data with clearinghouse data directly.

Table 1.6: Treatment Balance of Returning Sample—Clearinghouse Data

	Treatment	Control	Difference	P-value
Pre-treatment mean satisfaction with AI service provision (1-5)	4.212	4.248	-0.036	0.765
	[0.684]	[0.713]	(0.080)	
Pre-treatment mean veterinarian switching rate	0.047	0.026	0.020	0.131
	[0.218]	[0.206]	(0.019)	
Pre-treatment mean log AI visit charges	4.852	4.838	0.014	0.660
	[1.356]	[1.352]	(0.147)	
Pre-treatment mean AI success rate	0.694	0.669	0.025	0.541
	[0.445]	[0.439]	(0.051)	
Pre-treatment mean no. of cows	2.770	3.168	-0.398	0.351
	[2.785]	[2.349]	(0.384)	
Pre-treatment mean no. of buffalo	3.493	3.321	0.173	0.929
	[3.243]	[4.109]	(0.444)	
Pre-treatment mean distance to closest AI center (km)	2.413	2.007	0.406	0.728
	[2.158]	[2.190]	(0.245)	

Notes: Standard deviations reported in brackets. Standard errors reported in parentheses. Means and difference are unconditional. P-values are from OLS regressions with randomization strata fixed effects and standard errors clustered at the village-cluster. The sample consists of 300 farmer-level observations across 108 village-clusters of those farmers who received government AI service provisions both before and after receiving a treatment or control phone call. Some regressions have fewer observations due to missing data. Beginning in October 2014 treatment farmers received information about the AI success rates of their local government veterinarians. Satisfaction, AI visit charges, and numbers of cows and buffalo are reported by farmers on the phone one day after AI service provision. AI visit success rate is reported by farmers on the phone 60 days after AI service provision. Farmer switched vets and distance to closest AI center are automatically generated administrative data.

Table 1.7: Treatment Effects by Veterinarian Ranking—Clearinghouse Data

Outcome:	Returned	Switched veterinarians	Log price	AI success rate
	(1)	(2)	(3)	(4)
Panel A: Farmers told vet. was in top three in area				
Treatment farmer (=1)	0.008	-0.009	-0.169	0.010
	(0.013)	(0.035)	(0.136)	(0.115)
Mean of dependent variable	0.091	0.098	4.903	0.654
# Observations	1977	439	169	124
# Village-clusters	174	78	66	56
R-Squared	0.102	0.363	0.717	0.743
Panel B: Farmers told vet. was not in top three in area				
Treatment farmer (=1)	0.039*	0.005	-0.994	0.285*
	(0.020)	(0.079)	(1.419)	(0.161)
Mean of dependent variable	0.067	0.050	5.574	0.429
# Observations	1087	166	82	68
# Village-clusters	161	55	40	34
R-Squared	0.121	0.576	0.819	0.873
Sample	Pre	Post	Post	Post

Notes:*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors clustered at the village-cluster level reported in parentheses. All regressions include randomization strata fixed effects and controls for baseline mean outcomes (switched veterinarians, log price, AI success rate) as well as dummies for whether the given observation is missing each baseline mean outcome. In addition, columns (2) through (4) include exact call center script fixed effects and a time trend control. The sample for column (1) is farmers that received a government AI service and were subsequently treated, regardless of whether they then returned. The sample for columns (2) through (4) are farmers that returned after treatment. Note the differences in observations across columns are due to the fact that veterinarian switching can be detected without any successful phone calls, where as log price requires one successful phone call and AI success rate requires two successful phone calls to a farmer. Beginning in October 2014 treatment farmers received information about the AI success rates of their local government veterinarians. Returned is a dummy variable equal to one if a farmer that received government AI before treatment subsequently returned for government AI after treatment by the end of the project. Switched veterinarians is a dummy variable equal to one if the veterinarian that a farmer saw for a service provision was different than the last veterinarian seen. Log price is the log price paid for the service provision, as reported by the farmer when called to verify service provision. AI success rate is the rate of success of the AI services provided at a specific service provision upon follow up 60 days later. Panels are divided by whether a farmer was told when treated that his/her veterinarian from the last visit was in the top three or not, or would have been if s/he was not selected for control.

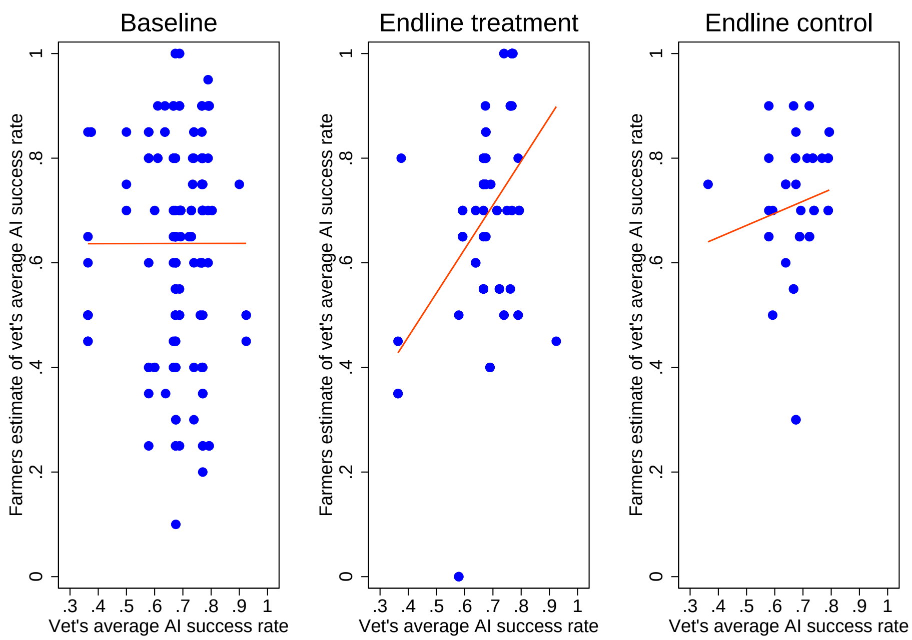

Figure 1.6: Treatment Effect on Farmer Expectations—Representative Survey Sample

Notes: The sample is farmers that received AI from a reported veterinarian that could be matched to our clearinghouse veterinarians. Farmer's estimates of vet's average AI success rate reported by farmers in baseline and endline surveys. Vet's actual average AI success rate is from clearinghouse data before October 2014. Beginning in October 2014 treatment farmers received information about the AI success rates of their local government veterinarians.

Table 1.8: Change in Farmer Expectations—Representative Survey Sample

	Farmer's estimate of vet's average AI success rate		
	(1)	(2)	(3)
Vet's actual average AI success rate	0.001	0.839**	0.231
	(0.177)	(0.385)	(0.229)
# Observations	145	66	37
# Village-clusters	34	21	20
R-Squared	0.000	0.162	0.020
Sample	Baseline	Endline T	Endline C

Notes :*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors clustered at the village-cluster level reported in parentheses. The sample is farmers that received AI from a reported veterinarian that could be matched to our clearinghouse veterinarians. Farmer's estimates of vet's average AI success rate reported by farmers in baseline and endline surveys. Column (1) limits to baseline responses by eventual treatment and control farmers. Column (2) limits to endline responses by treatment farmers. Column (3) limits to endline responses by control farmers. Vet's actual average AI success rate is from clearinghouse data before October 2014. Beginning in October 2014 treatment farmers received information about the AI success rates of their local government veterinarians. The null hypothesis that the coefficients in columns (2) and (3) are equal is rejected with a p-value of 0.115 from a regression interacting Vet's actual average AI success rate with a treatment indicator in the Endline sample.

Table 1.9: Treatment Effects by Farmer Expectations—Representative Survey Sample

Outcome:	Returned	Switched veterinarians	Log price	AI success rate
	(1)	(2)	(3)	(4)
Panel A: Farmers with above median expected-actual AI success				
Treatment farmer (=1)	-0.083	0.049	0.294	0.318
	(0.135)	(0.055)	(0.493)	(0.412)
Mean of dependent variable	0.370	0.231	5.688	0.500
# Observations	60	29	29	20
# Village-clusters	28	12	12	9
R-Squared	0.536	0.589	0.738	0.514
Panel B: Farmers with below median expected-actual AI success				
Treatment farmer (=1)	0.113	0.369	-1.399***	0.749*
	(0.274)	(0.329)	(0.385)	(0.370)
Mean of dependent variable	0.419	0.118	5.939	0.563
# Observations	53	32	28	28
# Village-clusters	29	16	14	16
R-Squared	0.468	0.756	0.898	0.588
Sample	Pre	Post	Post	Post

Notes:*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors clustered at the village-cluster level reported in parentheses. All regressions include randomization strata fixed effects and controls for baseline mean outcomes (switched veterinarians, log price, AI success rate) as well as dummies for whether the given observation is missing each baseline mean outcome. In addition, columns (2) through (4) include survey wave fixed effects and restricts the sample to those farmers that returned. The sample is limited to post treatment reports of AI service provision from farmers during our endline survey, conducted in June 2015. Returned is a dummy variable equal to one if a farmer that received government AI before treatment subsequently returned for government AI after treatment by the end of the project. Switched veterinarians is a dummy variable coded as one if the veterinarian a farmer saw for a service provision was different than the last veterinarian seen. Log price and AI success rates are recalled by farmers from service provisions two to seven months ago. Panels are divided above and below the median of veterinarian's farmers' estimate of their veterinarian's average AI success rate minus veterinarian's actual average AI success rate from clearinghouse data before October 2014. Positive values in this difference occur when farmers are told their veterinarian is better than they expected' negative values occur when farmers are told their veterinarian is worse than they expected. The median is .012.

1.8.2 Appendix Tables and Figures

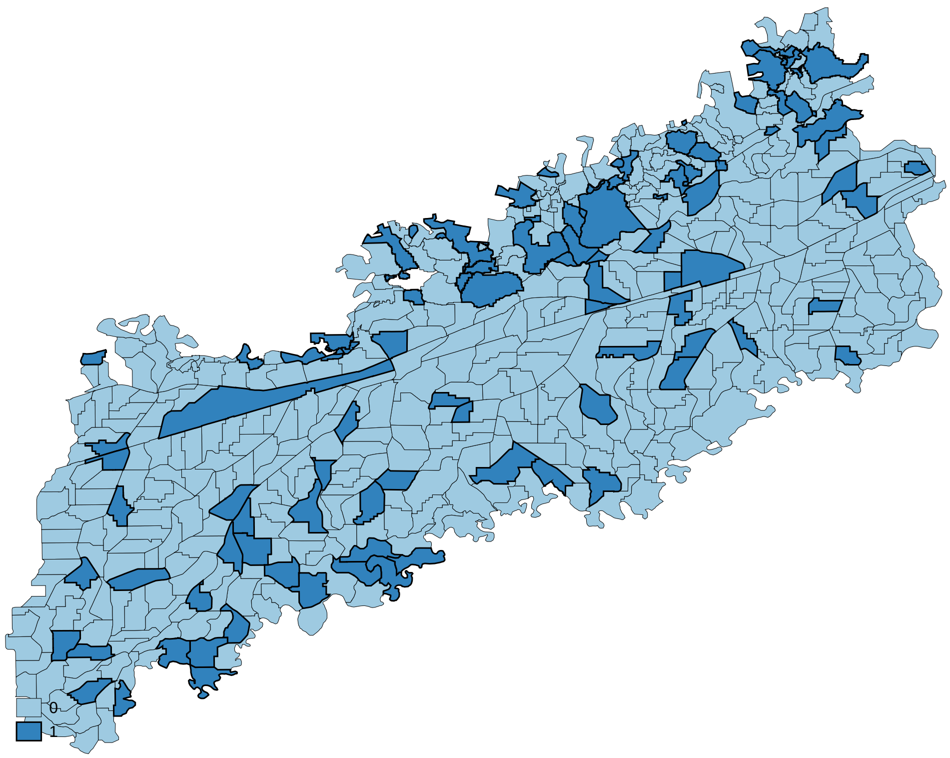

Figure 1.7: Representative Survey Sample Villages

Notes: Sampled villages are dark blue. The sample was stratified by whether or not a government veterinarian center was in the village and on whether the village was a canal colony. It is balanced along the following variables: area, settled area, cultivated area, area of wheat, rice, cotton, sugar cane, pulses, orchards, and vegetables, having a river, distance to the nearest veterinarian center, number of livestock in the village, literacy rates, religion, age, and standard wealth index characteristics. Results available upon request. Within each village, we selected ten households using the well-documented EPI cluster sampling method. In order to be surveyed, households had to report owning at least two livestock (cows and/or buffalo) and having regular access to a cellular phone.

Table 1.10: Treatment Balance—Representative Survey Sample, Additional Covariates

	Treatment	Control	Difference	P-value
Head of household education = None (=1)	0.388	0.404	-0.016	0.814
	[0.488]	[0.492]	(0.038)	
A child in the household attends public school (=1)	0.533	0.525	0.008	0.915
	[0.500]	[0.500]	(0.038)	
Household has used govt health services in past two years (=1)	0.399	0.466	-0.067	0.045
	[0.490]	[0.500]	(0.038)	
Amount of land household owns and rents for livestock	1.455	1.417	0.038	0.646
	[3.248]	[2.875]	(0.273)	
Household owns the house that they live in (=1)	0.926	0.948	-0.021	0.210
	[0.261]	[0.223]	(0.020)	
Hours of electricity per day	10.458	10.022	0.436	0.214
	[3.366]	[3.573]	(0.276)	
Household has a cooking stove/range (=1)	0.086	0.121	-0.035	0.119
	[0.280]	[0.326]	(0.024)	
Household made less than 100k PKR last year (=1)	0.320	0.301	0.019	0.349
	[0.468]	[0.460]	(0.036)	
Any member of household has hank account (=1)	0.235	0.275	-0.040	0.109
	[0.424]	[0.447]	(0.034)	
Believed it was likely that last vote was not secret (=1)	0.542	0.582	-0.040	0.396
	[0.499]	[0.494]	(0.041)	
Is likely to believe information given by gov't employee (=1)	0.776	0.815	-0.039	0.180
	[0.417]	[0.389]	(0.031)	
Average number of digits recalled	3.308	3.308	0.000	0.818
	[0.992]	[1.129]	(0.112)	
On a scale fo 0-10, how willing are you to take risks?	4.345	4.715	-0.370	0.332
	[3.008]	[6.894]	(0.503)	
Agreeableness	4.017	4.033	-0.016	0.756
	[0.743]	[0.702]	(0.057)	
Conscientiousness	4.071	4.128	-0.057	0.263
	[0.627]	[0.656]	(0.051)	
Extroversion	4.163	4.096	0.067	0.530
	[0.686]	[0.695]	(0.056)	
Neuroticism	2.363	2.375	-0.013	0.761
	[0.845]	[0.854]	(0.066)	
Openness	3.724	3.689	0.034	0.796
	[0.711]	[0.755]	(0.057)	

Notes: Standard deviations reported in brackets. Standard errors reported in parentheses. Means and differences are unconditional. P-values are from OLS regressions with randomization strata fixed effects and standard errors clustered at the village-cluster. The sample consists of 190 baseline farmer-level observations across 61 village-clusters. Some regressions have fewer observations due to missing data. All data come from baseline surveys fielded in August and September 2013. This sample of farmers was selected to be geographically representative of Sahiwal and is drawn from 90 different villages. The sample is limited to farmers that report receiving services from a government veterinarian at baseline. Treatment farmers received information about the AI success rates of their local government veterinarians. Treatment calls were conducted in November 2014 and January 2015. Agreeableness, conscientiousness, extroversion, neuroticism, and openness are all measures from the Big 5 Personality Index. These traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less).

Table 1.11: Treatment Effects—Representative Survey Sample

Outcome:	Returned	Switched veterinarians	Log price	AI success rate
	(1)	(2)	(3)	(4)
Treatment farmer (=1)	0.063	-0.058	0.027	0.470**
	(0.062)	(0.171)	(0.407)	(0.187)
Mean of dependent variable	0.222	0.152	5.852	0.581
# Observations	251	69	70	64
# Village-clusters	72	27	28	30
R-Squared	0.235	0.457	0.633	0.503
Sample	Pre	Post	Post	Post

Notes :*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors clustered at the village-cluster level reported in parentheses. All regressions include randomization strata fixed effects and controls for baseline mean outcomes (switched veterinarians, log price, AI success rate) as well as dummies for whether the given observation is missing each baseline mean outcome. In addition, columns (2) through (4) include survey wave fixed effects and restricts the sample to those farmers that returned. The sample is limited to post treatment reports of AI service provision from farmers during our endline survey, conducted in June 2015. Returned is a dummy variable equal to one if a farmer that received government AI before treatment subsequently returned for government AI after treatment by the end of the project. Switched veterinarians is a dummy variable coded as one if the veterinarian a farmer saw for a service provision was different than the last veterinarian seen. Log price and AI success rates are recalled by farmers from service provisions two to seven months ago.

Table 1.12: Comparing Farmers by Pre-treatment Veterinarian Ranking—Clearinghouse Data

	Vet. in top three	Vet. not top three
Satisfaction with AI service provision (1-5)	4.170	4.142
	[0.736]	[0.769]
Farmer switched vets since last AI visit	0.051	0.071
	[0.220]	[0.257]
AI visit charges (PKR)	192	212
	[170]	[269]
AI visit success rate (pregnancy / AI attempts)	0.628	0.635
	[0.477]	[0.476]
No of cows owned by farmer	2.382	2.668
	[3.154]	[3.660]
No of buffalo owned by farmer	2.816	3.516
	[3.165]	[5.949]
Distance to closest AI center (km)	1.710	3.257
	[1.572]	[2.949]

Notes: Standard deviations reported in brackets. The sample consists of 4,788 pre-treatment farmer-visit-level observations from 2,981 unique farmers that received government AI service provision. Some regressions have fewer observations due to missing data. Beginning in October 2014 treatment farmers received information about the AI success rates of their local government veterinarians. Satisfaction, AI visit charges, and numbers of cows and buffalo are reported by farmers on the phone one day after AI service provision. AI visit success rate is reported by farmers on the phone 60 days after AI service provision. Farmer switched vets and distance to closest AI center are automatically generated administrative data. Columns are divided by whether a farmer was told when treatment that his/her veterinarian from the last visit was in the top three or not, or would have been if s/he was not selected for control.

Chapter 2

Personalities and Public Sector Performance: Evidence from a Health Experiment in Pakistan

2.1 Introduction

Governments are the primary provider of services for the poor in developing countries. Yet, government employees, from front-line providers such as teachers and doctors to senior officials, commonly face weak incentives to perform (World Bank, 2004; Reinikka and Svensson, 2004; Chaudhury et al., 2006; Bandiera et al., 2009; Wild et al., 2012). A principal focus of many reforms aimed at improving service delivery is, therefore, to strengthen incentives.[1] Evidence supports the view that, in addition to incentives, personalitiy traits play a key role in determining performance (Borghans et al., 2008; Almlund et al., 2011; Heckman, 2011), can be changed (Kautz et al., 2014; Blattman et al., 2015), and that better recruitment policy can improve the personality profile of individuals selecting into public service (Dal Bó et al., 2013; Ashraf et al., 2014).[2] This suggests the possibility of strengthening services in developing countries through the separate avenue of personality traits.[3] This paper examines whether non-cognitive traits matter for public service delivery outcomes.

We consider three questions using a large-scale field experiment designed to

[1]Olken and Pande (2012) provide an overview of incentive reforms designed to reduce corruption and improve public sector performance more generally.

[2]Guided by insights from the field of industrial and organizational (I-O) psychology, firms, militaries, and governments in developed countries have long used psychometric measures to inform hiring, training, and promotion decisions. In a widely-cited meta-analysis of 85 years of data, for example, Schmidt and Hunter (1998) find that conscientiousness tests such as those in this paper not only predict job performance but do so while being much less correlated with general mental aptitude than years of education or job knowledge tests. Many others have stressed the predictive validity of these non-cognitive traits (Kaplan and Saccuzzo, 1997; Bowles et al., 2001; Heckman et al., 2006; Borghans et al., 2008; Groth-Marnat, 2009; Gatewood et al., 2010; Bazerman and Moore, 2012).

[3]Rasul and Rogger (2014) provide evidence that management practices are also an important determinant of public sector performance. In Nigeria, they find a strong positive relationship between a measure of managerial autonomy for bureaucrats and project completion, suggesting an additional means of improving service delivery beyond standard incentives.

improve health worker performance in Punjab, Pakistan.[4] First, do personality measures predict performance under status quo incentives, which are weak? Second, do these measures predict responses to a reform that changes incentives? Third, do these measures identify the senior officials who will react to information about the absence of their subordinates? This provides, to our knowledge, the first exploration of the relation between standard personality psychology measures and the performance of public sector officials. We examine these issues at three very different levels of the bureaucracy, from doctors working at the community level to senior administrators responsible for systems that serve several million people. The magnitudes we find in examining all three questions are systematically large and significant, supporting the view that this is a fruitful area for research.

In answer to our first question, the Big Five and Perry Public Service Motivation (PSM) measures systematically predict doctor and, to a lesser extent, inspector performance.[5] Doctors who score one standard deviation higher on the measured Big Five trait of conscientousness, for example, are 5.5 percentage points more likely to be present at work during an unanounced visit. Similarly, health inspectors that score one standard deviation higher on the measured PSM trait of commitment to policymaking are five percentage points less likely to be found colluding with doctors

[4]According to 2008 population estimates, Punjab is the ninth largest sub-national unit in the world with approximately 85 million citizens, of which 70 percent are rural. According to a 2011 report, the Punjab Department of Health provides outpatient services 90 percent of this total population per year, making it one of the largest health systems in existence. Despite the far reach of this system, Punjab performs poorly on major health indicators, with a infant mortality rate of 88 per 1000 live births, for example (National Institute of Population Studies, 2013).

[5]The Big Five personality traits, according to the Five Factor Model of personalities, are five separate dimensions of human personality that are thought to be highly descriptive and non-overlapping. These traits are agreeableness, emotional stability, extroversion, conscientiousness, and openness. The PSM measure is argued to capture attributes of individual personality relevant to the desire to provide public service. PSM has six traits—attraction to policymaking, commitment to policymaking, social justice, civic duty, compassion, and self-sacrifice.

to falsify inspection reports. In addition, health inspectors that score one standard deviation better on a proxy measure of the tendency to procrastinate are 5.8 percentage points more likely to complete each of their assigned inspections in a month.[6] Overall, we find significant positive correlations for four of eleven measured doctor personality traits and doctor attendance, and six of the remaining seven coefficients are also positive.[7] To help understand these coefficients, our back-of-the-envelope estimates indicate that replacing the bottom 25 percent of doctors in terms of the aggregate Big Five index with average doctors would result in about 4,650 or four percent more patients being seen by a doctor every month in Punjab's rural public health facilities.[8] A similar, though weaker, pattern holds with health inspectors, which we discuss in detail below.

To provide evidence on the second question, we designed and implemented a smartphone technology that verifies whether officials are performing regular facility inspections across Punjab, which we evaluated using a randomized control trial spanning the province.[9] We find that a one standard deviation increase in our mea-

[6]We obtain a proxy measure of an inspector's tendency to procrastinate by examining the degree to which the inspector tends to get his monthly quota of inspections done later in the month. Our approach is similar to that of Shapiro (2005) and Kuhn (2013), who use the steepness of the biweekly consumption profile to measure time preferences.

[7]Throughout the paper, we will scale our personality measures such that higher values are normatively better from the perspective of worker performance. We will report results both on individual traits and on summary indices of the Big Five and Public Sector Motivation traits. We acknowledge that focusing on indices rather than specific measures runs counter to the fundamental intention of these measures—to identify distinctive features of individual personality. We use indices for two reasons. First, we find that better traits predict better performance irrespective of the specific trait so that indices are useful for brevity. Second, from an econometric perspective, using an index provides a means of dealing with multiple hypothesis testing.

[8]If we focus on the specific Big Five personality trait of conscientiousness, consistently identified in the literature as highly relevant for performance, we find that replacing the bottom 25 percent of the distribution with doctors from the mean would result in 9,500 or 7.5 percent more patients being seen every month.

[9]Considering the distribution of personality types of agents most affected by an intervention can also help us understand what treatment effects might look like in other settings. On an intuitive level, if a bureaucracy is staffed with workers whose personalities are well-suited

sured aggregate Big Five index for a government inspector is associated with a 35 percentage point differential increase in inspections in response to treatment.[10]

On the final question, a one standard deviation increase above the mean in our measured aggregate Big Five index of a senior health official is associated with an additional 40 percentage point reduction in doctor absence at a facillity managed by the official when the facility's performance is experimentally flagged for the official's attention.[11] These officials oversee health systems responsible for several million citizens. The degree of our result suggests that improvements at this level of the bureaucracy might be particularly impactful.

The relationship between personality traits and policy outcomes in our data supports the recent focus on the selection and motivation of policy actors (Dal Bó et al., 2013; Ashraf et al., 2014, Forthcoming), the relationship between personalities and performance in other domains (Barrick and Mount, 1991; Salgado, 1997; Nyhus

to the job, increasing incentives to perform may make very little difference. Conversely, if workers are highly incompatible with their jobs, reforms may induce little additional effort. In line with this intuition, we find suggestive evidence that treatment effects from the monitoring technology are localized to the middle of the personality distribution.

[10]The only other variable that predicts a much higher treatment response is if the inspector has received a higher education degree, and both variables are significant predictors simultaneously.

[11]Both the results relating to the second and the third question are based on comparisons of treatment effects across different subgroups and so are not, themselves, experimentally identified (Deaton, 2010). Because personality is not randomly assigned, we can only argue that personalities strongly predict the types of individuals who will respond to changes in incentives. Relatedly, because we could potentially consider a number of different dimensions of heterogeneity, our statistical tests may not be of proper size (Miguel et al., 2014). We argue this should not be a major concern for three reasons. First, we designed the study expressly to understand the relevance of personality for performance. Other than checking staff attendance, we only collected data on the personalities and political connections of doctors, a dimension of heterogeneity we analyze in Callen et al. (2013). As evidence of this, we added an extra survey wave in which we tracked down doctors that we never found present in a clinic and in which we only measured personality traits at considerable effort and expense. Second, we composed a pre-analysis plan for this project in March of 2012, prior to the collection of any data on personalities. Finally, we also find similarly strong results even after correcting our standard errors using the method of Benjamini and Hochberg (1995).

and Pons, 2005; Heckman et al., 2006), and the potential malleability of personality traits (Kautz et al., 2014; Blattman et al., 2015). On selection into public service, Ashraf et al. (Forthcoming) find that both financial and non-financial incentives are more effective for more intrinsically motivated public health workers in Zambia and Ashraf et al. (2014), in the same context, find that health workers recruited by making career incentives salient perform better on the job than those recruited by making social incentives salient, despite being no less pro-social. Dal Bó et al. (2013) find that increasing wages substantially improves the pool of applicants to public jobs, as measured by IQ, Big Five, and Perry Public Sector Motivation.[12] Literatures in psychology and in economics also consistently point to a relationship between personality measures and economic success. For example, Heckman et al. (2006) find that measures of locus of control and self-esteem (traits related to neuroticism, one of the Big Five personality traits) from adolescence predict adult earnings to the same degree as cognitive ability. Similarly, Kautz et al. (2014) summarize a body of research finding that non-cognitive characteristics are often as predictive as cognitive skills in predicting economic success. Nyhus and Pons (2005) find using Dutch household data that wages are correlated with two of the Big Five personality traits, emotional stability and conscientiousness.[13] Other meta-analyses find conscientiousness to be consistently predictive of earnings (Barrick and Mount, 1991; Salgado, 1997). Hogan and Holland (2003) find in a meta-analysis that all five Big Five measures positively predict performance on specific job criteria, and that the predictions become stronger as the job criteria become more specific.[14] Regarding whether traits are fixed, Kautz

[12]Our results directly complement this paper as we find that workers with higher scores on the Big Five and Public Sector Motivation measures work more often and more effectively in a similar context with weak extrinsic incentives. Taken together, this suggests that increasing wages can improve service delivery by causing more effective workers to select into public service.

[13]These two traits are also the most consistently predictive of performance in our data.

[14]There is also more general evidence that the traits of senior executives are important in

et al. (2014), in a comprehensive review of the literature, argue that the evidence so consistently supports malleability that non-cognitive attributes should be called "skills", rather than "traits", partly to re-orient policy toward the value of investing in these dimensions of human capital.[15]

These three literatures, combined with the positive relationship between better traits and better performance in our data, suggest three respective ways that taking non-cognitive attributes into consideration can improve service delivery. First, the finding that the psychological profile of applicants to public jobs can be affected by the recruitment process suggests delivery outcomes can be improved via selection. Second, given broad evidence that traits are malleable, delivery could be improved by measures that strengthen non-cognitive attributes. Third, psychometric measures might be useful as diagnostics in hiring or promotion decisions.[16] The degree of correlation between personality measures, doctor attendance, and the responsiveness of senior officials complements these literatures by showing that improving the stock of non-cognitive skills in the public sector can translate into better service delivery

determining the performance of the entities that they manage. At the firm level, Johnson et al. (1985) find that shareholder wealth is positively correlated with measures of a firm's executive's 'talents' and 'decision-making responsibility.' Bertrand and Schoar (2003) find that a significant extent of the heterogeneity in investment, financial, and organizational practices of firms can be explained by the presence of manager fixed effects. Malmendier et al. (2011) find that overconfidence affects management decisions. At the cross-national level, Jones and Olken (2005) find, using deaths of leaders as exogenous variation, that leaders matter for a country's growth, especially when constraints on the executive are weak.

[15]Similarly, Roberts et al. (2006) examine 92 studies for patterns in the mean-level of Big Five personality traits. The authors find that people increase in measures of social dominance (a facet of extroversion), conscientiousness, and emotional stability as they age, especially over ages 20 to 40. Blattman et al. (2015) find in an experiment that providing Cognitive Behavioral Therapy (CBT) to high-risk Liberian men caused their conscientiousness scores and other measures of self-control to improve after just eight weeks. It is important to note that the psychological literature is in agreement, however, that these measured personality traits are more than situational specific, and thus are worthwhile to use for explanatory purposes as we do in this paper (Roberts, 2009).

[16]Klinger et al. (2013) discuss the merits and disadvantages of using psychometrics to screen for loan provision. A major concern, which applies equally in the public sector, is the potential for strategic misrepresentation of personality type.

outcomes.

While our data allow us to make some progress on relating personalities to performance, they also face some limitations. First, because our sample includes officials in positions of power, obtaining measures of cognitive ability was thought to be potentially demeaning. We therefore are unable to directly compare the relevance of cognitive and non-cognitive attributes for service delivery. Second, as in much of the literature, no component of the personality traits we measure is exogenously determined, limiting our ability to identify the causal relationship between personalities and performance. To address this, in our information experiment with senior officials, we aimed to manipulate a factor affecting performance—information about the performance of their suborndinates—that most plausibly should be mediated through the mechanism of personalities.

The paper proceeds as follows: Section 2.2 outlines a simple model based on Almlund et al. (2011) to explain how personality traits can affect job task selection and performance. Section 3.2 provides the institutional details necessary to understand our results. Section 2.4 outlines our research design and reports results. Section 3.6 concludes.

2.2 Framework

In this section, we provide a framework to help us understand the first two questions considered in this paper—do personality measures (i) predict performance under status quo incentives and (ii) predict responses to a reform that changes incentives?[17]

Let our personality measures capture a worker's type, θ, with cumulative dis-

[17] A number of papers incorporate personality traits into standard economic models such as the Roy Model (Almlund et al., 2011) or the principle-agent framework (Besley and Ghatak, 2005; Benabou and Tirole, 2003).

tribution $F(\theta)$. Let performance be the binary decision that a doctor or health inspector makes of whether to attend work. If a worker attends, he receives a fixed salary of W and incurs a cost of effort $\lambda(\theta)$. If a worker shirks, he exerts no effort and receives the fixed salary with probability $1-p$ and an arbitrarily small punishment c with probability p, as well as an outside option of Q.[18]

2.2.1 Personality Type and Performance

The marginal worker indifferent between working and shirking will satisfy

$$W - \lambda(\theta) = (1-p)W - pc + Q. \tag{2.1}$$

If work is less costly for better types ($\frac{\partial \lambda}{\partial \theta} < 0$), then all workers with θ greater than that of the marginal worker will choose to work. Equation 2.1 therefore gives that workers with better personality types are weakly more likely to attend work. This accords with Almlund et al. (2011), in which the authors define traits as features which allow individuals to produce more with a fixed amount of effort.[19]

2.2.2 Personality Type and Responses to Changes to Incentives

We now turn to predictions regarding how changes to incentives affect the decision to work. Consider a worker of type θ_m who is just indifferent between working and shirking. To see what happens when the probability of detection p changes, note

[18] We choose Q here to denote 'quack', the term in Pakistan for a private doctor. We use the 'he' pronoun because almost all government doctors and health inspectors are men.

[19] This might be because workers with better personality types are more efficient with their time or because the psychic costs required to achieve a given task are lower. Or, in a simple utility framework, we can think of θ as the ratio of the marginal utility from work to the marginal utility from leisure for a worker.

that

$$\theta^M = \lambda^{-1}(p(W + c) - Q) \tag{2.2}$$

$$\frac{\partial\theta^M}{\partial p} = \frac{1}{\lambda'(\lambda^{-1}(p(W + c) - Q)}. \tag{2.3}$$

Given our earlier assumption that $\frac{\partial\lambda}{\partial\theta} < 0$, and assuming that $p(W + c) > Q$, it must be that $\frac{\partial\theta^M}{\partial p} < 0$, or that the marginal worker's personality type decreases with an increase in detection probability.

We can see this in a simple picture in Figure 2.1. Let θ^{M1} be the marginal worker before an increase in p and θ^{M2} the lower-type marginal worker afterwards. All workers with $\theta > \theta^{M1}$ continue to work and workers with types in the shaded area $\theta^{M1} > \theta > \theta^{M2}$ are induced to work by the increase in detection probability. The types induced to work are the highest (best) among those that shirk prior to the shift in p. Equation 2.1 therefore also describes how a personality type relates to a reform in incentives.[20]

Here we assume personality traits only affect the cost of effort in an otherwise simple indifference equation. It follows that better personality types are more likely to work ex-ante and that the better types among ex-ante shirkers will be more likely to respond to an increase in incentives. The decision to work is potentially much more complex. For example, personality traits that are useful in the public sector may also increase productivity in the outside option (i.e., Q may also be a function of θ).

[20]Note that Figure 2.1 allows us to make two additional points. The first is that the results in this paper, as with all results from randomized interventions, are Local Average Treatment Effects. That is, our intervention may induce some workers to work, but there are some workers that will always work and some that will never work regardless of the intervention. The second point is that the initial position of θ^{M1} matters significantly to the size of the impact of an increase in detection probability. This also highlights the importance of the shape of the distribution of personality types, as a very narrow distribution might see different effects than a uniform distribution from an increase in p. Both the initial position of θ^{M1} and the distribution of personality types can be estimated ex-ante, allowing for better targeted policies.

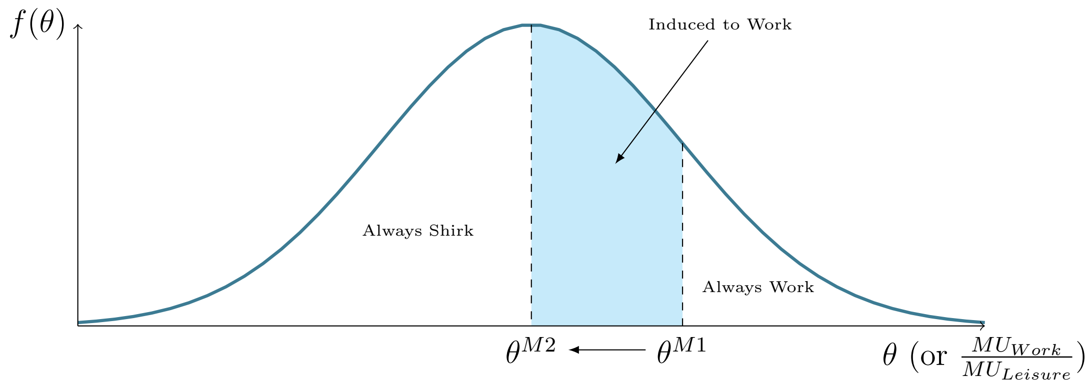

Figure 2.1: Effect of an Increase in Detection Probability on the Decision to Work or Shirk

More generally, θ might not only capture the single-dimensional productivity gains to personality traits. It may also capture heterogeneity in workers' outside options, in workers' cognitive ability, in workers' ability to mitigate political pressure from outside their office, and so on. We could deal with this in two ways. Most simply, we could redefine $\lambda(\theta)$ to include the all of these personality trait-dependent costs and benefits. Then the simple model would encapsulate a richer understanding of these costs and benefits of personality traits, but it would be unable to differentiate these costs and benefits. Second, we could enrich the model by, for example, modeling Q as a function of θ. Without additional and somewhat implausible assumptions, doing so immediately expands the set of predictions to the point where the model is no longer falsifiable. We demonstrate this in Appendix Section 2.7.3.

2.3 Public Health Services in Punjab

This section describes the main institutional details relevant to our experiment and our empirical results.

In Punjab, the provision of health care services is managed by the Department of Health. Authority in the department is highly centralized in the upper echelons of the bureaucratic hierarchy. Senior actors described in this section play a central role in determining the quality of delivery. They are also responsible for a substantial number of facilities spread, in many cases, across vast geographic distances. This presents a major challenge for monitoring that we aim to address with our smartphone monitoring system.

The main performance outcomes in this paper are measured at primary front-line public health clinics, called Basic Health Units (BHUs).[21] BHUs are designed to be the first stop for rural patients seeking medical treatment in government facilities, providing mainly primary services, including out-patient services, neo-natal and reproductive health care, and vaccinations against diseases. Hereafter in this paper, we use the word 'clinic' interchangeably to describe BHUs. There are 2,496 BHUs in Punjab.[22] Almost all BHUs are located in rural and peri-urban areas. Each facility is headed by a doctor, known as the Medical Officer, who is supported by a Dispenser, a Lady Health Visitor, a School Health and Nutrition Supervisor, a Health/Medical Technician, a Mid-wife and other ancillary staff. Officially, clinics are open, and all staff are supposed to be present, from 8AM to 2PM and patients seen in these clinics are required to pay a nominal fee of around $0.01 USD per visit.

2.3.1 Health Sector Administration

Figure 2.2 depicts a simplified version of the health administration hierarchy in Punjab. District governments are responsible for managing local health facilities.

[21]There are five major types of facilities: (i) Basic Health Unit (BHU); (ii) Rural Health Center (RHC); (iii) Tehsil Headquarter Hospital (THQ); (iv) District Headquarter Hospital (DHQ); and (v) Teaching Hospital. In Punjab, a tehsil is the largest administrative sub-division of a district. There are 121 tehsils across 37 districts.

[22]Each Basic Health Unit serves approximately one Union Council (Union Councils are smallest administrative units in Pakistan).

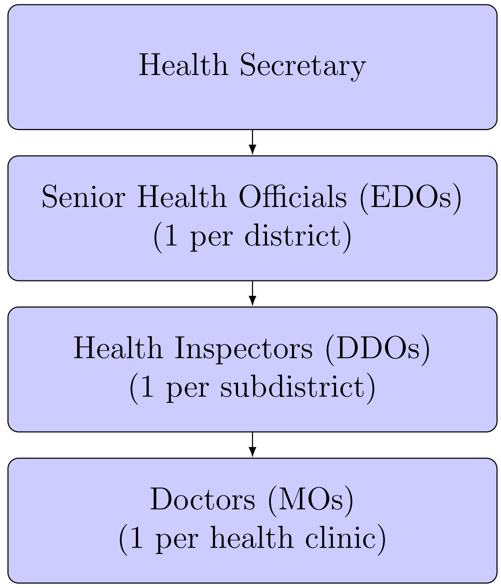

Figure 2.2: Health Sector Administration in Punjab

Each District Department of Health is headed by an Executive District Officer (EDO) who reports both to the official in charge of the district and to two provincial health officials.[23] EDOs are directly supported by several Deputy District Officers (DDOs). DDOs primarily inspect and manage health facilities.[24] DDOs are required to inspect every clinic in their jurisdiction at least once a month and record information collected during the visit on a standard form. DDOs have the authority to punish a clinic's absent staff by issuing a formal reprimand, suspending staff, and/or withholding pay (in the case of contract staff). Each Medical Officer is similarly responsible for their own clinic, with proportional duties. Throughout the paper, we will refer to Executive District Officers as senior health officials, Deputy District Officers as inspectors, and

[23]The senior official in charge of the district is the District Coordinating Officer (DCO). The provincial health officials are the Director General of Health Services and the Secretary of the Department of Health.

[24]While inspections are the primary official functions of the DDO, our time use data indicate that, on average, DDOs spend 38.9 percent of their time on inspections and management, with the remainder of their time principally spent managing immunization drives. For full details please see Callen et al. (2013).

Medical Officers as doctors, focusing on their role rather than their title.

As is true in many developing countries, low health worker attendance is a major issue in Punjab. From unannounced visits to clinics in 2011, we find that only 56 percent of clinics were inspected in the prior two months, and that doctors were only present 43 percent of the time when one was posted.[25] This points to a lack of enforcement that allows health inspectors and doctors to shirk. In the next section, we provide results related to the role of personality traits in the performance of senior officials, inspectors, and doctors.

2.4 Results

In this section, we present three sets of results, each corresponding to one of the three questions laid out in Section 3.1. First, we study correlations between the measured personality traits of doctors and health inspectors, their job performance (attendance and inspections respectively), and their propensity to collude with one another. Second, we use these measures to predict health inspectors' response to an experimental intervention which increases the probability of detecting shirking. Finally, we examine whether traits identify which senior health officials react to information about the absence of their subordinates. This analysis relies on a second policy intervention which manipulates the information provided to senior officials about the absence of their subordinates.

2.4.1 Predicting Doctor and Health Inspector Performance

We first examine if personality measures predict bureaucratic performance under status quo incentives, for doctors and then for health inspectors. We measured

[25] Doctors were not posted at 35 percent of clinics, which means unconditional doctor presence was only 32 percent.

personality for 389 doctors in Punjab posted to a representative sample of 850 of the 2,496 rural health clinics in Punjab.[26] We also measured personality for the universe of health inspectors and senior health officials in Punjab, for a total of 101 inspectors and 33 senior health officials. For all 850 clinics in our sample, we also measured attendance during unannounced visits in November 2011, June 2012, and October 2012. Before presenting results, we describe the monitoring intervention that allowed us to collect our data as well as our personality, procrastination, and performance measures.[27]

Measuring Personality

The personality measurement batteries in this paper are from personality psychology and are used broadly, including recently in economics. We use two measures: the Big Five personality traits and the Perry Public Service Motivation (PSM) traits.

Developed by psychologists in the 1980s, the Five Factor Model is now one of the most widely used personality taxonomies in the field.[28] We measure the Big Five traits using a 60 question survey developed specifically in Urdu and validated for use in Pakistan by the National Institute of Psychology at Quaid-i-Azam University, Islamabad. Each of the 60 questions offers the respondent a statement such as "I see myself as someone who does a thorough job" and asks them to agree or disagree with the statement on a five point Likert scale (Disagree strongly, Disagree a little, Neutral, Agree a little, or Agree strongly).[29] In addition to measuring Big Five

[26] 306 facilities in this sample have no doctor posted. We omit these clinics from our doctor analyses.

[27] We interviewed inspectors and officials through pre-arranged office visits. To account for frequent doctor absence and transfers, we interviewed doctors in two unannounced independent inspections, followed by pre-scheduled interviews. We succeeded in interviewing 389 of an estimated 544 posted doctors, or 72 percent of our sample population.

[28] See John et al. (2008) for a summary of the measures and its history. Borghans et al. (2008) provide a summary of empirical results in psychology and economics.

[29] John et al. (2008) provide the mapping between questions and traits.

traits separately as the mean response to twelve questions (where disagree strongly is assigned a 1, disagree a little a 2, etc.), all traits are normalized into z-scores and averaged to form a single Big Five index.[30]

The Perry Public Service Motivation (PSM) battery is designed to measure intrinsic motivation for public service. Also developed in the 1980s, it comprises a total of 40 questions measuring six traits—attraction to policymaking, commitment to policymaking, social justice, civic duty, compassion, and self-sacrifice.[31]

Table 2.2 reports summary statistics for these measures separately for doctors and health inspectors in treatment and control districts in our randomized control evaluation of a new monitoring technology.[32] There is substantial variation in personality traits across individuals consistent with the original intention of the battery: to capture substantial and important differences in personality types.[33] It is this heterogeneity that allows for the possibility of linking differences in personality to variation in performance.

[30]The results presented in the following sections are robust to a 'naive' personality index in which each of the 60 questions is individually normalized and then one average z-score is formed. These results are available on request. The psychology literature emphasizes the distinctions between these five traits. While important and, in our application, useful for considering mechanisms, these distinctions are not necessary for simple prediction. In order to facilitate understanding such distinctions, however, we always present the index alongside trait-by-trait results. We discuss the potential relevance of specific traits for policymaking in the conclusion.

[31]Perry and Wise (1990) and Perry (1996) introduce the battery and Petrovsky (2009) provides a summary of studies using this measure.

[32]We describe the experiment in Subsection 2.4.1 below. We capture these measures after treatment is administered. However, balance on these measures increases our confidence that they are stable over the horizon of the study and unaffected by treatment. The full distributions for these measures are reported in Table 2.6. Summary statistics for senior health officials are reported in Table 2.7.

[33]Borghans et al. (2008) explain the development of the Big Five.

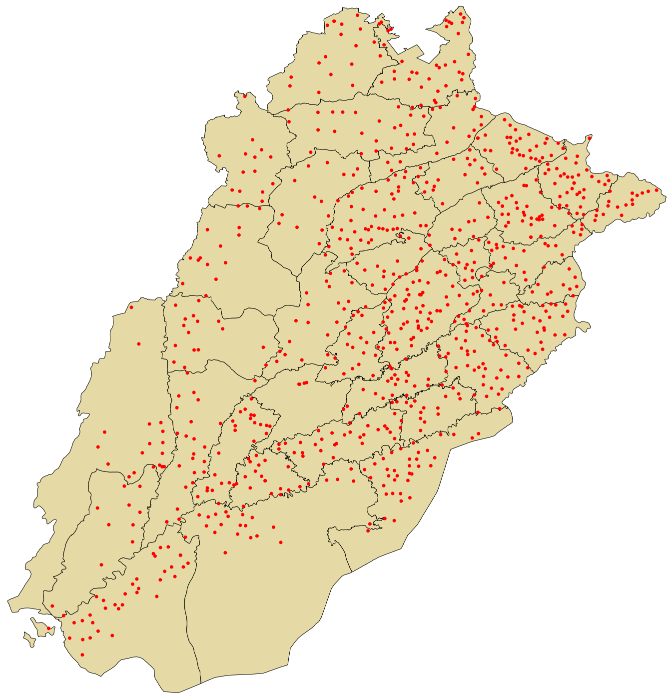

Figure 2.3: Locations of Clinics (Basic Health Units) in the Experimental Sample

Measuring Doctor Performance

To obtain measures of performance, we collected primary data on a representative sample of 850 of the 2,496 clinics or Basic Health Units in Punjab. Clinics were selected randomly using an Equal Probability of Selection design, stratified on district and distance between the district headquarters and the clinic. Our estimates of absence are, therefore, self-weighting and require no sampling correction. All districts in Punjab except Khanewal—the technology pilot district—are represented in our data. Figure 2.3 provides a map of clinics in our experimental sample along with the district boundaries in Punjab.

Information on staff absence, health inspections, and facility usage was collected through three independent and unannounced inspections of these clinics. We visited each facility three times: November 2011, June 2012, and October 2012. Our survey team interviewed and physically verified the presence of the Medical Officer, or doctor, and verified the last health inspection that occurred through written records stored at the facility.[34]

We have two measures of doctor job performance: (i) whether doctors were present during our unannounced visits, and (ii) a proxy measure of collusion between doctors and health inspectors to falsify reports. We define collusion as a dummy variable coded as one when the doctor is absent during both of our post-treatment unannounced visits and is marked present during every single health inspection during the treatment period.[35] Baseline performance measures for doctors are reported in Table 2.6.

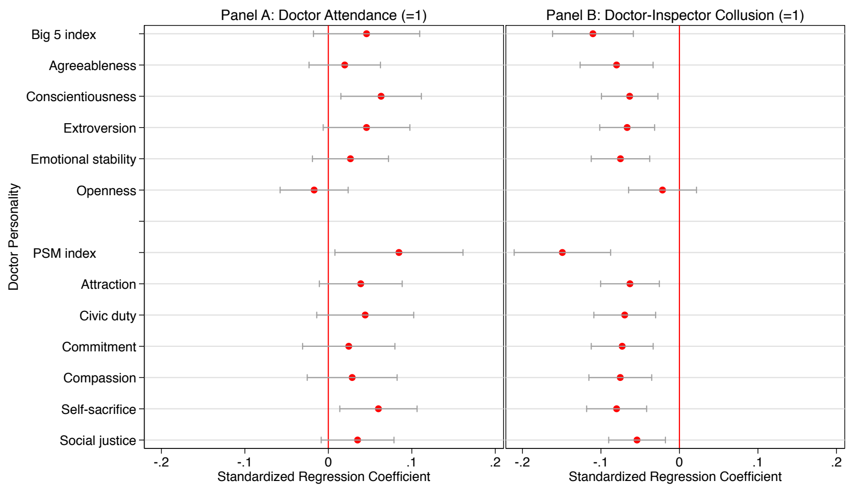

Figure 2.4: Personality and Performance: Doctors

Notes: Each regression coefficient reported comes from a separate regression of the performance measure, Doctor Attendance in Panel A and Doctor-Inspector Collusion in Panel B, on the indicated doctor personality measure. Error bars represent 95 percent confidence intervals, with standard errors clustered at the clinic level. All regressions include Tehsil (county) and survey wave fixed effects. In all cases, personality measures are normalized to have mean zero and standard deviation of one in the sample, and thus the regression coefficients reported can be interpreted as the impact of a one standard deviation increase in a given personality trait or aggregate measure. The sample is restricted to control district clinics for which doctor personality data are available and a doctor is posted. Regressions corresponding to the figure are reported in Appendix Tables 2.8 and 2.9.

Personality and Doctor Performance

In Figure 2.4, Panel A shows that doctors that score one standard deviation above the mean on the Big Five measure of conscientiousness are about five percentage points more likely to be present at work during an unannounced visit. Similarly, two measures of PSM, civic duty and self-sacrifice, are also significantly predictive. Finally, all but one coefficient are positively correlated with doctor attendance. In Panel B, we find that doctor personality measures are even stronger predictors of collusion between health inspectors and doctors. Doctors who score one standard deviation higher on measured civic duty, for example, are about 6 percentage points more likely to be identified as potentially colluding. Ten out of eleven Big Five and PSM traits are highly predictive of collusion, with negative signs.[36]

We draw two lessons from this exercise. First, in Appendix Table 2.10, we

[34]In addition, the attendance of Dispensers, Health/Medical Technicians, Lady Health Visitors, Midwives, and School Health and Nutrition Specialists were also recorded. Survey teams were trained at regional hubs (four in total) by senior enumerator trainers and our team members. Following these trainings, the teams made visits to clinics in their assigned districts and remained in regular contact with their team leaders and our research team. Surveys took three weeks to field for each wave. The attendance sheet for the staff was filled out at the end of the interviews and in private. Inspectors record visits by signing paper registers maintained at the health facility. We measure whether an inspection occurred by interviewing facility staff and verifying the register record. Data collection and entry followed back-checks and other validation processes consistent with academic best practice.

[35]The median number of health inspections for each facility in our treatment sample is 12, with a max of 50. The collusion we have in mind occurs when a health inspector calls a doctor before an inspection to alert him to be in attendance. Then, after the health inspector records his presence, the doctor is under very little pressure to attend until he gets another similar phone call from the inspector. Of course, such patterns in the data could arise by chance, though the chance decreases with the number of inspections. As such, we have run all of our collusion analysis using weighted least squares and we find results very similar to those OLS results presented below. Results provided upon request. The strong correlation we find between these measures and personality types also suggests that the proxy is successfully capturing malfeasance. An immediate problem with this proxy is that it partly reflects attendance. We deal with this by adjusting our inference to reflect multiple hypothesis tests using the False Discovery Rate procedure described in Benjamini and Hochberg (1995).

[36]See Appendix Tables 2.8 and 2.9 for point-estimates.

find that personality is a stronger predictor for doctors than three other plausibly important observables—doctor tenure in the department of health, doctor tenure at the specific health clinic at which the doctor worked at the time of the survey, and the distance from this clinic to the doctor's hometown in Pakistan (in travel time). Though we have only a limited number of covariates for this exercise, they are potentially correlated with a wide number of factors relevant to the relationship between personality and performance. Overall tenure, for example, will be correlated with age, experience, and the number of relationships with others in the health department. Tenure at a specific facility will be correlated with how much influence a doctor has in the Department of Health as transfers are frequent and often undesirable. Distance to home might proxy for the desirability of a posting as in interviews doctors frequently expressed a strong desire to work near their home and family.

Second, the degree of the estimated coefficients is meaningful. As one example, we look at the relationship between doctor attendance and out-patient services. To do so, we first establish a strong positive correlation between doctor presence at their clinic during one of our unannounced visits and reported out-patients seen at that clinic in Appendix Table 2.11. Given this correlation and those in Figure 2.4, in Appendix Section 2.7.4 we perform a simple back-of-the-envelope exercise to quantify the effect we would expect on out-patients seen if we were to replace the bottom 25 percent of doctors with average doctors in terms of the Big Five index and one specific trait—conscientiousness. We find an expected increase of 4,646 and 9,469 out-patients seen per month from replacing the bottom 25 percent of doctors with average doctors according to the Big Five index and the conscientiousness measure respectively.[37]

[37]We treat this exercise speculatively for several reasons. Importantly, it relies on an assumption that the correlation we observe reflects a causal linkage between personality traits and performance. However, in combination with the substantial literature suggesting that personality matters for performance and the related literature on teacher value added

Monitoring Intervention

We collected personality data during a larger experimental policy reform that considered audits by government monitors as a solution to the problem of bureaucratic absence. The "Monitoring the Monitors" program replaced the traditional paper-based monitoring system for clinic utilization, resource availability, and worker absence with an android-based smartphone application. In the new system, data generated by health inspections are transmitted to a central database using General Packet Radio Service (GPRS). Data are then aggregated and summary statistics, charts, and graphs are presented in a format designed in collaboration with senior health officials to effectively communicate information on health facility performance. These data are also: (i) geo-tagged, time-stamped, and complemented with facility staff photos to check for reliability; and (ii) available in real time to district and provincial officials through an online dashboard. The objective of this monitoring system is to make the activities of health inspectors available to their senior officials in real time. Figure 2.5 shows one view of the online dashboard.[38]

In the context of our model above, providing data to senior officials creates a discrete increase to p, the probability that a health inspector will be caught if he is failing to do his inspections. Prior to Monitoring the Monitors, and in control districts, the paper-based monitoring system severely limits a senior officials ability to monitor inspectors. In treatment districts, on the other hand, reports are immediately and automatically sent up the chain of command, and the required geo-tags, time stamps, and photos serve as instant verification that the inspector and all reported staff are

(Chetty et al. (2014a); Chetty et al. (2014b); Hanushek (2011); Hanushek and Rivkin (2012); Staiger and Rockoff (2010)), it does not seem too outlandish to argue that performance would improve substantially if the distribution of types were improved.

[38]Application development started in August 2011. After developing the application and linking it to a beta version of the online dashboard, the system was piloted in the district of Khanewal. We remove Khanewal district from the experimental sample. Health administration staff were provided with smartphones and trained to use the application.

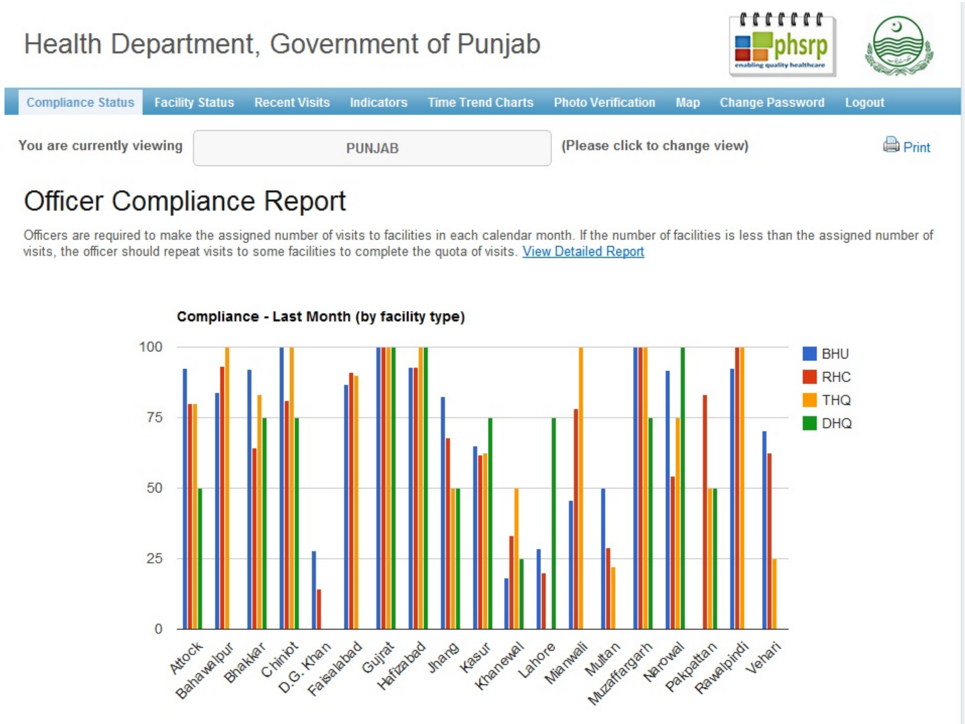

Figure 2.5: Online Dashboard—Summary of Inspection Compliance by District

present at the clinic being inspected.[39]

Measuring the Tendency to Procrastinate

A nascent literature uses intertemporal consumption and effort profiles to measure time preference and time inconsistency.[40] Inspectors in Punjab are required to inspect every facility in their jurisdiction once a month. The intertemporal inspection allocations captured by our smartphone monitoring system reveal patterns indicating a tendency to procrastinate for a majority of our inspectors.

[39]See Callen et al. (2013) for the core results from the broad Monitoring the Monitors experiment.

[40]Augenblick et al. (Forthcoming) elicit time preferences based on the intertemporal allocation of non-monetary tasks in the lab. Shapiro (2005) and Kuhn (2013) provide evidence that the intra-month consumption profile of food stamp recipients reflects dynamically-inconsistent planning and better fits a quasi-hyperbolic model than a standard exponential discounting model.

Panel A of Figure 2.6 depicts the average number of inspections on each day of the month. On the first day of the month, inspectors perform an average of about 0.31 inspections. After the first ten days of the month, average inspections on a given day are roughly 0.8. The time profile of inspections over a month has a positive slope. Several months of data allow estimation of the slope of the intertemporal profile of inspections, providing a proxy measure of each inspector's tendency to procrastinate. We estimate

$$Inspections_{d,m} = \alpha + \eta \ Day \ of \ Month_{d,m} + \delta_m + \epsilon_{d,m} \tag{2.4}$$

where $inspections_{d,m}$ is the number of inspections on a given day d in a month m, δ_m are fixed effects for each month, and $Day \ of \ Month$ runs from one to 28 depending on the calendar day of the month.[41] Inspectors with a positive η estimate do fewer inspections at the beginning of the month and more at the end as they approach the deadline for their quota, suggesting a tendency to procrastinate.

Panel B of Figure 2.6 provides a histogram of the estimates of η for 36 inspectors. 29 of these 36 inspectors have positive slope coefficients. The average slope coefficient is 0.014, which indicates that over the course of the month the number of inspections per day increases by about 0.4.

[41] The effective deadline for inspections is the 28th of the month as senior officials and inspectors meet during the final days of the month to review the month's inspections. We only include months for which we have complete information for a health inspector and drop holidays. We retain data for 36 health inspectors and have an average of 8.75 months of inspection-level data per inspector. The median number of inspections in a month is 25 and inspectors are responsible for between 4 and 46 facilities with a median of fifteen. Two factors limit our sample. First, we only have daily inspection data for treatment districts, which include roughly 50 health inspectors. Of these inspectors, we drop fourteen who transferred into treatment districts taking over the phone of the previous inspector. Transfer records do not indicate the date of transfer, making it impossible to identify the period of smartphone data that correctly corresponds to these 14 inspectors.

Panel A: Inspections by Day of Month

Panel B: Histogram of Slope Parameters

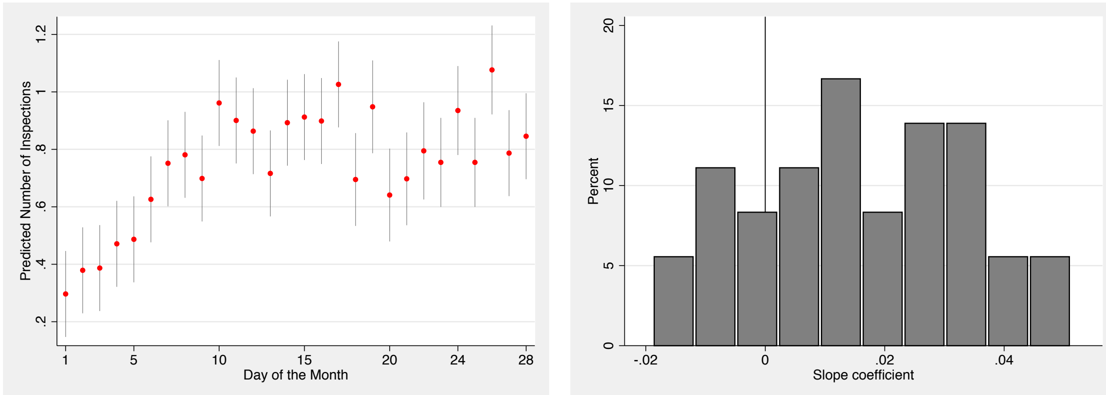

Figure 2.6: The Temporal Allocation of Inspections

Notes: Panel A plots the predicted number of inspections from a regression of inspections on dummies for each day of the month, and for each month, as well as a control for the number of facilities in the inspector's jurisdiction. Panel B is a histogram of slope parameters obtained from estimating Equation (2.4) separately for each of the 36 vaccinators in our sample.

Measuring Inspector Performance

We have two measures of job performance for health inspectors: (i) a dummy equal to one if the facility records an inspection in the two months prior to an unannounced visit; and (ii) the same proxy measure of collusion between doctors and health inspectors to falsify reports as described in Section 2.4.1. These measures were obtained during the same three independent and unannounced inspections of health clinics described in Section 2.4.1. Baseline performance measures for health inspectors are reported in Table 2.6.

Procrastination and Inspector Performance

As with our personality measures, we can correlate our proxy measure of the tendency to procrastinate with health inspector performance. In Table 2.1, we present results of a regression of health inspections on our estimated time slope coefficient.

We see that health inspectors with larger time slope coefficents (reflecting a larger tendency to procrastinate) conduct fewer inspections, once you limit the sample to those inspectors with at least nine facilities in their jurisdiction (the 10th percentile in terms of health facilities per district across the sample). Specifically, we see that a one standard deviation increase in the procrastination measure is associated with a 6.7 percentage point decrease in the probability that an inspection was carried out in the last two months at a health clinic. This relationship may reflect a limitation on the number of inspections that can be carried out in a fixed period of time. Those who delay all of their inspections until the end of the month are not able to complete their monthly assignment.

Personality and Inspector Performance

We examine how much the personalities of health inspectors predict their job performance in control districts (i.e., those under status quo incentives) in Figure 2.7. In Panel A, we consider the relation between personalities and whether an inspection was carried out in the last two months. In Panel B, we see that PSM traits are associated with less collusion, enough to distinguish the coefficient on the aggregate z-score from zero. In this case, health inspectors that score one standard deviation higher on aggregate PSM are about seven percentage points less likely to be identified as potentially colluding.[42]

[42]See Appendix Tables 2.12 and 2.13 for complete details on the results summarized in Figure 2.7. The estimates in Figure 2.7 indicate a negative relationship between both conscientiousness and emotional stability and the number of inspections. These coefficients both reflect $p < 0.10$ and suggest that better traits are associated with worse performance. These coefficients are estimated only on the subsample of 298 clinics in control districts which have a doctor posted. In Appendix Tables 2.14 and 2.15, we find no evidence of a correlation on the full sample of 424 control facilities, indicating that inspectors with better traits are more likely to have inspected facilities *without* doctors posted. There is therefore some weak evidence that better inspectors substitute away from better facilities with a doctor posted toward more rural facilities without a doctor.

Table 2.1: Procrastination and Inspector Performance

	Health Inspection in Last Two Months (=1)				
	(1)	(2)	(3)	(4)	(5)
Time Slope Coef. (Standardized)	-0.001	-0.060*	-0.067*	-0.079**	-0.060*
	(0.041)	(0.024)	(0.027)	(0.027)	(0.022)
Mean of dependent variable	0.708	0.695	0.723	0.723	0.723
# Observations	456	420	357	357	357
# Tehsils	32	28	25	25	25
R-Squared	0.221	0.242	0.241	0.249	0.256
Inspector Jurisdiction Size Percentile:	0	10	25	25	25
Controls for Big Five Traits	NO	NO	NO	YES	NO
Controls for PSM Traits	NO	NO	NO	NO	YES

Notes: This table reports on the correlation between an inspectors tendency to procrastinate and their inspection performance. Column 1 provides estimates from an OLS regression of a dummy equal to one if a facility was inspected in the last two months on the time slope coefficient. The time slope coefficient is estimated for each inspector using a regression of the number of inspections done on a given day of the month on a day of the month variable, with month fixed effects. We then standardize the variable across inspectors. Higher time slope coefficients indicate a larger tendency to procrastinate. Standard errors clustered at the Tehsil (sub-district) level—the jurisdiction of a given inspector—are reported in parentheses. All regressions include district and survey wave fixed effects. The sample is limited to health inspectors in treatment districts for which we have daily inspection data. The 10th percentile # Health Clinics in an inspectors Tehsil corresponds to 9 clinics, the 25th percentile to 12 clinics. The median number of health clinics in a Tehsil is 19 and the max is 46. Controls for Big Five Traits include agreeableness, conscientiousness, extroversion, emotional stability, and openness. Controls for PSM traits include attraction to policymaking, commitment to policymaking, social justice, civic duty, compassion, and self-sacrifice. *Levels of Significance:* $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

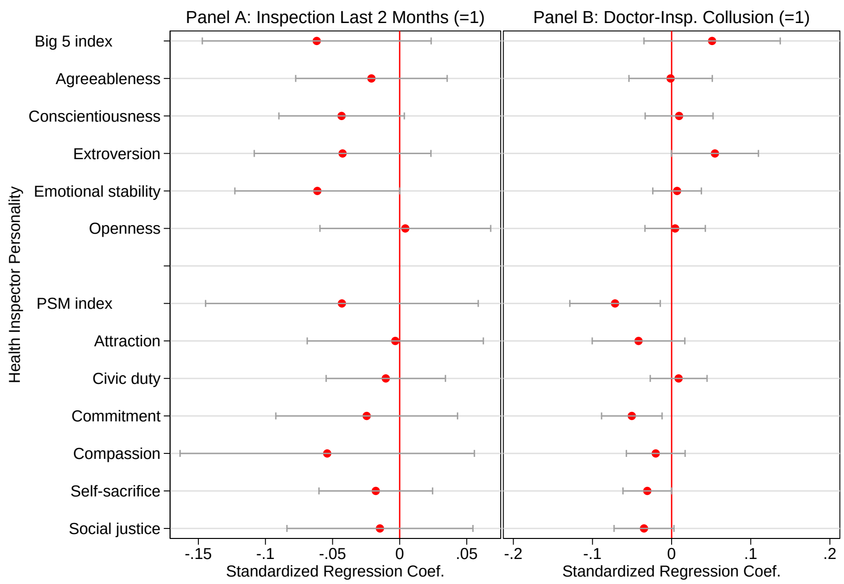

Figure 2.7: Personality and Performance: Health Inspectors

Notes: Each regression coefficient reported comes from a separate regression of the displayed performance measure on the indicated standardized health inspector personality measure. Error bars represent 95 percent confidence intervals. Standard errors are clustered at the clinic level. All regressions include Tehsil (county) and survey wave fixed effects. In all cases, personality measures are normalized to have mean zero and standard deviation of one in the sample, and thus the regression coefficients reported can be interpreted as the impact of a one standard deviation increase in a given personality trait or aggregate measure. The sample is restricted to control district clinics for which doctor personality data are available and a doctor is posted. Appendix Tables 2.12 and 2.13 provide corresponding regression tables.

2.4.2 Personalities and Treatment Response Heterogeneity

We now consider whether personality traits, including the tendency to procrastinate, predict health inspectors' response to a reform that increased incentives to complete inspections.

Evaluating the Smartphone Monitoring

Our experimental sample comprises all health facilities in the district of Punjab, which has a population of at least 85 million citizens. Tens of millions of public sector health users therefore were potentially affected by the program. As described above, we monitored a subsample of 850 clinics, drawn to be representative of facilities in the province, using independent and unannounced inspections.[43] We randomly implemented the program in 18 of the 35 districts in our experimental sample. In assigning treatment, we stratified on baseline attendance and the number of clinics in a district to ensure a roughly even number of treatments and controls. Figure 2.8 depicts control and treatment districts.[44]

Personality and Treatment Response

We investigate whether impacts of the monitoring program are heterogeneous by the personality type of the inspector. Table 2.2 presents personality measures by treatment status for doctors and health inspectors. There is one significant difference

[43]These are the same clinics and inspections from the correlations presented in the previous section.

[44]Treatment is randomized at the district level. The intervention channels information about inspections to district health officials; a design randomizing treatment at an administrative unit beneath the district, say the tehsil, would very like result in treatment affecting control units. The Department of Health also viewed sub-district randomization as not administratively feasible. Cluster randomization also allays some concerns about externalities generated by interactions between inspectors in the same district. All inspectors in a district are required to attend monthly meetings. While they typically have frequent interactions within districts, these relations are almost non-existent across districts.

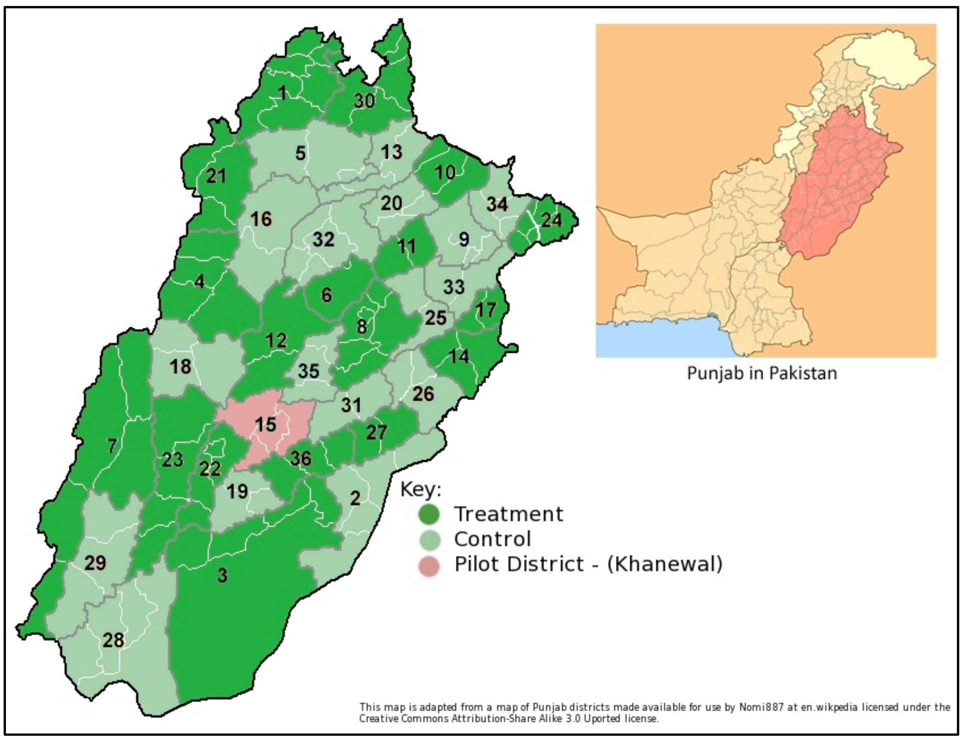

Figure 2.8: Treatment and Control Districts

in the balance table—treated health inspectors have slightly lower civic duty scores than those in control groups on average. This is plausibly due to sampling fluctuation as it is a fairly small difference and the only one among the 27 differences estimated.

We consider the effects of an increase in health inspector monitoring on their performance by inspector personality. Results are presented in Table 2.3.[45] We estimate regressions using the difference-in-difference specification

$$Y_{dit} = \beta_0 + \beta_1 Trait_{di} + \beta_2 Treatment_{dit} + \beta_3 Treatment_{dit} x Trait_i + \delta_t + \lambda_i + \varepsilon_{dit} \quad (2.5)$$

where Y_{dit} is a dummy equal to one if a facility records an inspection in the prior two months, $Treatment_{dit}$ is a variable equal to one for treated districts during the post-treatment periods (waves two and three), where i refers to the clinic, d refers to the district, and t to the survey wave, and $Trait_i$ is a personality trait of the inspector overseeing facility i. δ_t and λ_i are survey wave and clinic fixed effects, respectively. We cluster all standard errors at the district level.

For health inspectors, there are heterogeneous effects of our experiment on the rate of health inspections. Health inspectors with a Big Five index one standard deviation above the mean, for example, exhibit a 35 percentage point higher treatment effect in terms of health inspections. With an unconditional mean inspection rate of 66 percent, inspectors with a z-score one standard deviation above the mean come very close to completing all of their inspections as a result of treatment. We decompose this effect in columns (5)-(9) and find that that it is being driven most strongly by emotional stability—the trait of being able to capably respond to new stressors and

[45] Our other previous measure of performance, collusion between inspectors and doctors, cannot be studied in this context because the construction of collusion relies on data from our treatment districts' smartphone app. We have no information on health inspector-reported doctor attendance in the control districts of the Monitoring the Monitors experiment.

Table 2.2: Treatment Balance on Doctor and Health Inspector Personality

	Big Five Personality Traits							
	Doctor Personality Traits				Inspector Personality Traits			
	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Big Five Index	-0.058	0.042	-0.100	0.295	-0.017	0.018	-0.035	0.802
	[0.713]	[0.820]	(0.095)		[0.637]	[0.745]	(0.140)	
Agreeableness	3.498	3.577	-0.079	0.309	3.783	3.659	0.124	0.231
	[0.622]	[0.678]	(0.077)		[0.477]	[0.541]	(0.103)	
Conscientiousness	3.958	3.996	-0.037	0.605	4.159	4.117	0.041	0.679
	[0.548]	[0.570]	(0.072)		[0.452]	[0.536]	(0.100)	
Extroversion	3.624	3.686	-0.062	0.277	3.703	3.734	-0.031	0.754
	[0.464]	[0.501]	(0.057)		[0.525]	[0.459]	(0.099)	
Emotional Stability	-2.647	-2.536	-0.111	0.180	-2.461	-2.338	-0.124	0.307
	[0.641]	[0.702]	(0.082)		[0.571]	[0.624]	(0.120)	
Openness	2.926	2.932	-0.006	0.907	3.020	3.113	-0.093	0.264
	[0.372]	[0.451]	(0.050)		[0.471]	[0.350]	(0.083)	
	Perry Public Sector Motivation							
	Doctor Personality Traits				Inspector Personality Traits			
	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
PSM Index	-0.017	-0.018	0.001	0.989	-0.061	0.071	-0.131	0.288
	[0.695]	[0.691]	(0.079)		[0.621]	[0.614]	(0.123)	
Attraction	3.481	3.442	0.039	0.581	3.552	3.568	-0.016	0.881
	[0.630]	[0.610]	(0.070)		[0.532]	[0.568]	(0.110)	
Civic duty	4.182	4.184	-0.002	0.969	4.255	4.435	-0.180	0.034
	[0.594]	[0.526]	(0.059)		[0.415]	[0.424]	(0.084)	
Commitment	3.773	3.774	-0.001	0.982	3.915	3.969	-0.054	0.514
	[0.511]	[0.463]	(0.050)		[0.458]	[0.370]	(0.083)	
Compassion	3.493	3.546	-0.053	0.432	3.743	3.659	0.085	0.380
	[0.515]	[0.516]	(0.067)		[0.475]	[0.488]	(0.096)	
Self Sacrifice	4.065	4.080	-0.015	0.820	4.316	4.395	-0.079	0.396
	[0.563]	[0.574]	(0.065)		[0.482]	[0.454]	(0.093)	
Social Justice	3.950	3.906	0.044	0.464	4.098	4.200	-0.102	0.268
	[0.571]	[0.619]	(0.060)		[0.490]	[0.430]	(0.092)	
# Health Workers	242	147			51	48		

Notes: Variable standard deviations reported in brackets. Standard errors clustered at the district level reported in parentheses. The doctor sample is limited to clinics where a doctor is posted at baseline. The Big Five and PSM traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively. Actual observations for each regression vary by a small amount based on no responses.

Table 2.3: Testing for Heterogeneous Impacts of Monitoring by Personality Type

	Health Inspection in Last Two Months (=1)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: Big Five Personality Traits									
Monitoring (=1)		0.178	0.022	-0.006	0.010	0.003	0.030	-0.033	0.023
		(0.154)	(0.129)	(0.114)	(0.109)	(0.115)	(0.124)	(0.118)	(0.129)
Monitoring x Big Five Index				0.351**					
				(0.133)					
Monitoring x Agreeableness					0.170*				
					(0.094)				
Monitoring x Conscientiousness						0.186*			
						(0.102)			
Monitoring x Extroversion							0.116		
							(0.098)		
Monitoring x Emotional Stability								0.210**	
								(0.083)	
Monitoring x Openness									0.195
									(0.126)
Mean of Dependent Variable		0.642	0.656	0.656	0.656	0.656	0.656	0.656	0.656
# Observations		1331	1145	1145	1145	1145	1145	1145	1145
# Clinics		644	547	547	547	547	547	547	547
R-Squared		0.048	0.048	0.069	0.069	0.062	0.053	0.064	0.063
PANEL B: Public Service Motivation									
Monitoring (=1)	0.178	0.033	0.023	0.026	0.039	0.024	0.012	0.041	0.021
	(0.154)	(0.126)	(0.120)	(0.111)	(0.127)	(0.111)	(0.119)	(0.130)	(0.122)
Monitoring x PSM Index			0.202						
			(0.140)						
Monitoring x Attraction				0.211**					
				(0.078)					
Monitoring x Civic duty					-0.029				
					(0.066)				
Monitoring x Commitment						0.103			
						(0.082)			
Monitoring x Compassion							0.184		
							(0.115)		
Monitoring x Self Sacrifice								0.016	
								(0.090)	
Monitoring x Social Justice									0.014
									(0.102)
Mean of Dependent Variable	0.642	0.649	0.649	0.649	0.649	0.649	0.649	0.649	0.649
# Observations	1331	1164	1164	1164	1164	1164	1164	1164	1164
# Clinics	644	555	555	555	555	555	555	555	555
R-Squared	0.048	0.051	0.057	0.076	0.051	0.062	0.062	0.054	0.053

Notes: This table reports heterogeneous impacts of our smartphone monitoring treatment by personality type. Column (1) reports average treatment effects on treatment and control district clinics. Columns (2) - (10) are limited to clinics in tehsils for which health inspector personality data is available. The difference in observations between Panels A and B is due to one inspector answering the PSM but not the Big Five survey. The Big Five and PSM traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). All personality traits are then normalized across inspectors. The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively. *Levels of Significance:* $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

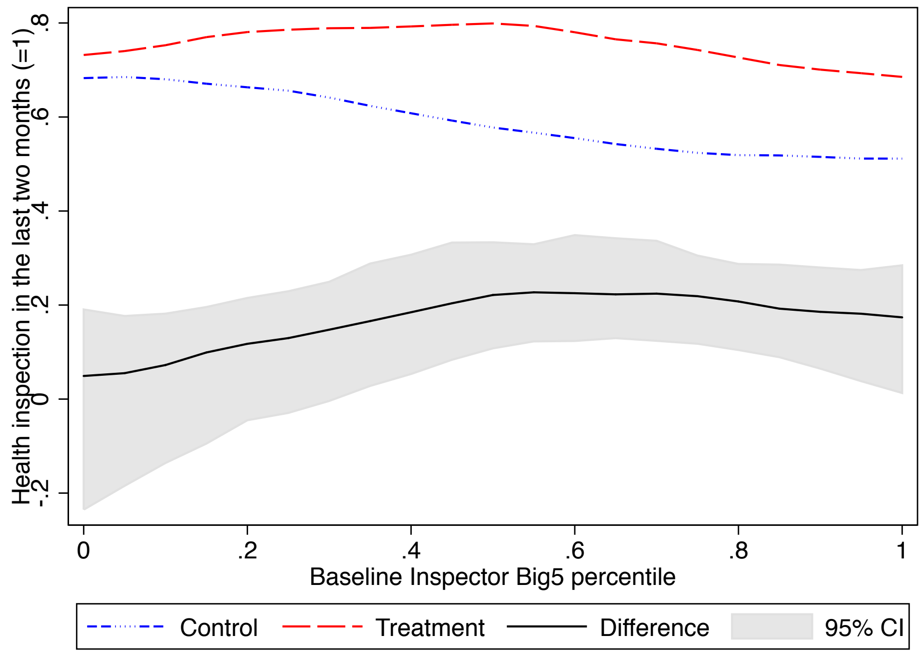

Figure 2.9: Nonparametric Treatment Effects

This figure plots a kernel-weighted local polynomial regression of whether a clinic had a health inspection in the last two months on every 5th percentile of baseline Big Five index separately for treatment and control districts, as well as the difference at each 5th percentile of baseline scores. The confidence intervals of the treatment effects are constructed by drawing 1,000 bootstrap samples of data that preserve the within-district correlation structure in the original data and plotting the 95 percent range for the treatment effect at each 5th percentile of baseline scores.

demands. Besides openness, all Big Five traits have positive and large coefficients. We also see some positive and similarly large effects of the PSM index, attraction, and compassion within the PSM traits, though only attraction is significant.[46]

[46] Note that to test for robustness in our effects to the small number of district clusters in our analysis, we have conducted Fisher exact tests for all results. In all cases, the estimated p-value is as at least as significant as from OLS. We have also separated the differential effects into our two post-treatment survey waves and find that the results sustain over time for as long as we were able to follow health clinics (roughly one year after treatment began). This is important because in Callen et al. (2013), we document that the overall treatment effects on health inspections do in fact fade by the second survey wave. Results available upon request.

Figure 2.9 presents nonparametric treatment effects of health inspector Big Five index across the distribution of inspectors according to the Big Five summary measure. We can see that the effect in Table 2.3 is primarily being driven by those health inspectors in the middle of the Big Five distribution. This fits the extended model presented in Section 2.7.3 in which it is plausible that the effects of this intervention are localized to those inspectors in the middle of the distribution. See Appendix Figures 2.11 and 2.12 for nonparametric treatment effects trait-by-trait. While the location of the treatment effect peaks varies by trait, the overall shape is similar for specific traits.[47]

There are two more points to make about these experimental results. First, as you can see in Appendix Table 2.17, personality does at least as much to predict the response to increased monitoring as all of the other covariates that we record for health inspectors. Completion of higher education is also a consistent predictor, but it predicts separately from personality. Second, these correlations are of a meaningful magnitude. Increased inspections may not lead to an overall increase in doctor attendance, but they generate information that is helpful in the case that a health inspector or more likely a senior health official *is* interested in enforcing attendance. We will see this directly in the next subsection.

2.4.3 Predicting Response to Information

In this section, we examine whether personality identifies the senior health officials who will react to information about the absence of their subordinates. To do this we study the response of senior officials, as measured by doctor absenteeism in clinics under their supervision, to a second policy intervention in which we manipulated the presentation of information to these officials.

[47]Note that the point estimates in Figure 2.9 do not match those from Table 2.3. This is due to the fact that the regressions in the table include survey wave and clinic fixed effects.

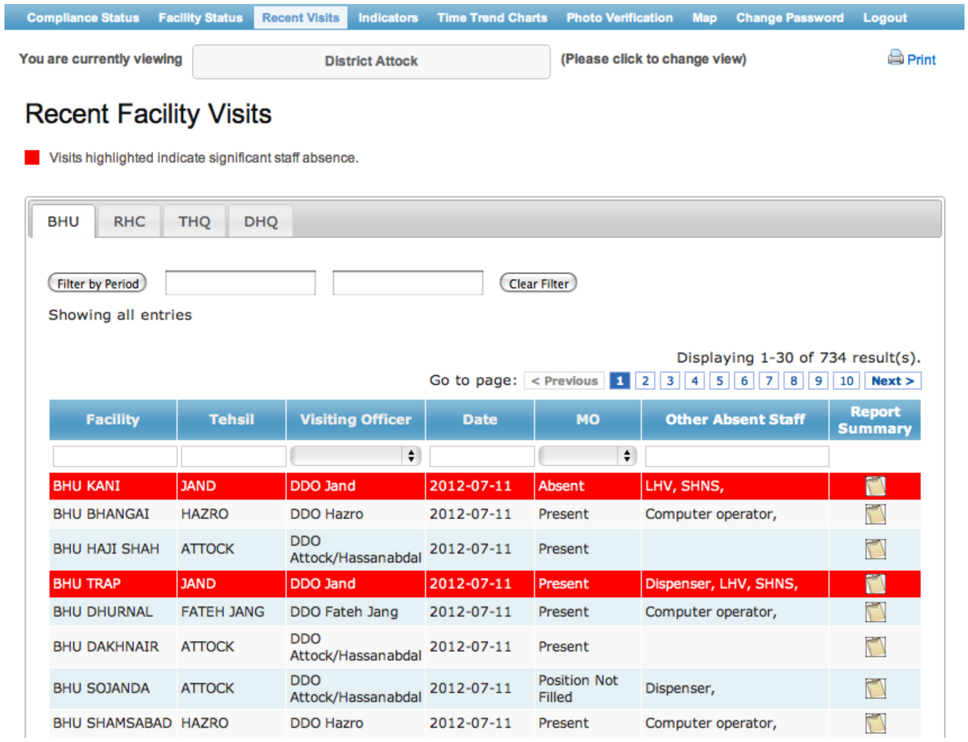

Figure 2.10: Highlighting Underperforming Facilities to Test Mechanisms

Information Experiment

The Monitoring the Monitors system aggregates data from health inspections and presents them to senior health officials in each district of Punjab on an online dashboard. This dashboard is only visible to these senior health officials as well as to the Secretary of Health for Punjab and the Director General of Health for Punjab. Figure 2.10 provides an example of a dashboard view visible to senior health officials.

To test whether senior health officials react to information about the absence of their subordinates, we directly manipulated the data on the dashboard to make certain underperforming facilities salient. This was achieved by highlighting in red, or "flagging" reports by inspectors that found three or more staff absent at a clinic.[48]

[48] In Callen et al. (2013), we examine at length whether this manipulation affects subsequent doctor absence, finding consistent evidence that flagging facilities leads to decreased subsequent doctor absence.

This cutoff of three or more staff absences was set by our research team and was not communicated to any of the doctors, health inspectors, or senior health officials. We selected this cut-off based on the distribution of staff absence from baseline data. The peak of the distribution lies at two or three absent staff, suggesting that a cut-off at the center of this peak would yield the highest power to detect an effect of flagging in red.

Though the cutoff was purposefully arbitrary, our motivation for making absence data salient was not. Senior health officials in Punjab are in charge of health service provision in their district. These officials are constantly receiving information from facilities, staff, and citizens. Given the volume of information available to these officials, we designed the intervention to test whether making information salient could catalyze action by senior health officers.

Personality Predicts Response to Information

Appendix Table 2.7 presents summary statistics for senior health officials in Punjab, which are similar in magnitude to summary statistics of both doctors and health inspectors. We examine whether manipulating attendance information affects subsequent doctor absence with the following specification

$$Absent\ Survey_{it} = \psi_0 + \psi_1 Trait_i + \psi_2 Flagged_{it-1} + \psi_3 Trait_i * Flagged_{it-1} + \delta_t + \eta_{it} \tag{2.6}$$

where $Absent\ Survey_{jt}$ is equal to one if the doctor posted to facility i was absent during our unannounced visit in wave t, $flagged_{it-1}$ is a dummy equal to one if the facility was flagged in red on the dashboard prior to survey wave t, $Trait_i$ is a personality measure for the senior official in charge of facility i, $Absent\ Dashboard_{it-1}$ is equal to one if the doctor was noted as absent in the period prior to our survey during the official inspection, and δ_t are survey wave fixed effects.

Table 2.4: Heterogeneity in the Information Treatment by Senior Official Personality

	Doctor Present (=1)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: Big Five Personality Traits									
Clinic Flagged as Underperforming on Dashboard		0.161*	0.146	0.159	0.140	0.144	0.132	0.154	0.163
		(0.095)	(0.103)	(0.098)	(0.103)	(0.100)	(0.105)	(0.100)	(0.110)
Flagged x Big Five Index				0.402**					
				(0.200)					
Flagged x Agreeableness					0.086				
					(0.144)				
Flagged x Conscientiousness						0.172*			
						(0.097)			
Flagged x Extroversion							0.097		
							(0.096)		
Flagged x Emotional Stability								0.185*	
								(0.105)	
Flagged x Openness									0.051
									(0.106)
Mean of Dependent Variable		0.563	0.520	0.520	0.520	0.520	0.520	0.520	0.520
# Observations		142	123	123	123	123	123	123	123
# Clinics		122	106	106	106	106	106	106	106
R-Squared		0.226	0.204	0.231	0.206	0.227	0.211	0.219	0.205
PANEL B: Public Service Motivation									
Clinic Flagged as Underperforming on Dashboard	0.161*	0.146	0.165	0.146	0.155	0.254**	0.153	0.146	0.201*
	(0.095)	(0.103)	(0.105)	(0.103)	(0.104)	(0.121)	(0.110)	(0.103)	(0.108)
Flagged x PSM Index			0.124						
			(0.169)						
Flagged x Attraction				0.072					
				(0.102)					
Flagged x Civic Duty					0.027				
					(0.089)				
Flagged x Commitment						0.231			
						(0.148)			
Flagged x Compassion							-0.028		
							(0.114)		
Flagged x Self Sacrifice								-0.032	
								(0.100)	
Flagged x Social Justice									0.139
									(0.097)
Mean of Dependent Variable	0.563	0.520	0.520	0.520	0.520	0.520	0.520	0.520	0.520
# Observations	142	123	123	123	123	123	123	123	123
# Clinics	122	106	106	106	106	106	106	106	106
R-Squared	0.226	0.204	0.208	0.207	0.204	0.217	0.204	0.204	0.219

Notes: This table tests for heterogeneity in the impact of providing information about underperforming clinics to senior officials by the personality types of the senior officials. Clinics were flagged as underperforming if three or more of the seven staff were absent in one or more health inspections of the clinic fifteen to fourty-five days prior to an unanounced visit by our survey enumerators. All columns restrict the sample to those clinics where only two or three staff were absent (up to seven staff can be marked absent). In addition, the sample is limited to Monitoring the Monitor treatment districts due to the necessity of the web dashboard for flagging clinics. Column (1) reports un-interacted impacts of flagging. Columns (2) - (10) are further limited to clinics in districts for which senior health official personality data is available.Standard errors clustered at the clinic level reported in parentheses. All regressions include district and survey wave fixed effects and condition on a doctor being posted. *Levels of Significance:* $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Facilities are flagged only if three or more staff members are absent. Consequently, if we restrict our sample to only facilities where, in the month prior to our unannounced visit, only two or three staff were absent, we can estimate the effect of flagging on a sample where the only difference might plausibly be whether the facility was flagged.[49]

Table 2.4 reports results from this test, limiting the sample to facilities with two or three staff absent during an inspection. Facilities flagged as underperforming to a senior official with a Big Five z-score one standard deviation above the mean subsequently experience an increase in doctor attendance that is 40 percentage points greater than a facility flagged to a senior official at the mean Big Five index.[50]

There are several ways through which the above effect may have operated. For instance, the health officials could have taken formal action against delinquent workers, or they could simply have censured the officers informally. While we are unable to discern this effect given our data, anecdotally, we have learned that the second channel is more likely to work, given limited powers for hiring and firing people.

Appendix Table 2.20 provides suggestive evidence that senior health officials with higher personality types stepped up the share of their time spent monitoring

[49]In Appendix Table 2.18 we verify the drop in absence for people who score higher on the Big Five index is limited to right around the discontinuity, with a waning, though significant, effect in a slightly larger window.

[50]Note that in Table 2.4 we cannot reject the null hypothesis that the interaction term on the Big Five index is different than the uninteracted flagging effect. In Appendix Tables 2.19, we show that when senior health officials' are split into quartiles by Big Five index, we can significantly reject that those in the bottom and top quartile have the same flagging effect (with a substantial differential effect). We define the window during which a clinic can be flagged in red prior to one of our unannounced visits as 15 to 45 days before our visit. Senior health officials only looked at the web dashboard every week or two, so we would not expect an immediate response from flagging. However, if the window is made too long, virtually every facility will become flagged and we will lose variation. The p-values of the significance of the coefficient on the Big Five index and PSM index for a wide range of windows are reported in Appendix Figures 2.13 and 2.14. These figures also indicate that we have not selected the window most favorable for our result.

health facilities in response to dashboard flags. You can see senior health officials with a one standard deviation higher Big Five index increased the share of their time spent monitoring health facilities by 3.1 percentage points for each facility that was flagged in their district in the window prior to our collection of their time use information (wave three). The mean number of flags per district in this time-frame was 7.88, which translates to large increases in time spent monitoring by better personality types in response to flags. Although, this evidence is at best suggestive because it is based on seventeen observations.

The worry with the above results is that senior health officials might be substituting other work with increased monitoring of health facilities. The data suggest that senior health officials may have decreased their share of time spent on the lunch prayer break, on work related to monthly polio vaccination drives, and on 'other work' in response to flags. Unfortunately, these effects are not significant individually.[51]

As with the correlational and experimental results above, we show that personality is a better predictor of the response to information than other important covariates for senior health officials. See Appendix Table 2.21 for these results.

The results presented in this section provide another validation of personality measures in predicting performance, this time in the case of senior health officials. Personality measures predict which senior health officials will react to information about the absence of their subordinates with large magnitudes. Simply flagging high absence clinics in red essentially eliminates doctor absence in clinics overseen by senior health officials one standard deviation above the mean in terms of their Big Five index. These results also speak to potential mechanisms. It seems plausible that the same information treatment provided to individuals in highly comparable positions results in different real world impacts because different personality types

[51] Category-by-category time use tables available by request.

take different action in response to information.

Turning to impact, if we perform an exercise similar to the exercise for inspectors above, and replace the bottom eight senior health officials in terms of measured Big Five index with the average individual, with 27.3 percent of facilities flagged we would expect to see an increase of 10.8 percent in the number of monthly outpatient visits, an increase of 12,598 visits per month. Thus replacing eight senior health officials could have a larger impact than we would expect from replacing over 600 doctors.

2.4.4 Summary of Results

This paper performs several tests of associations between personalities and objective performance measures for public health workers at three different levels of the bureaucracy in Punjab, Pakistan. These results are summarized in Table 2.5.[52]

We run regressions reflecting seven separate tests for eleven different traits, for a total of 77 hypothesis tests. Of these, 30 regressions return statistically significant coefficients in the predicted direction at the 90 percent significance level. If we adjust standard errors using the False Discovery Rate procedure in Benjamini and Hochberg (1995) to account for multiple hypothesis testing, thirteen regressions return statistically significant coefficients in the predicted direction at the 90 percent significance level.

2.5 Conclusion

Governments, like any organization, are made of people with potentially stark interpersonal differences. We find that these differences are useful in predicting per-

[52]Appendix Table 2.22 presents the same table but with standard errors adjusted using the False Discovery Rate procedure in Benjamini and Hochberg (1995).

Table 2.5: Results Summary

Alternative Hypothesis:	Personality Predicts Performance				Personality Predicts Monitoring Treatment Heterogeneity	Personality Predicts Information Treatment Heterogeneity
Public Actor:	Doctor		Inspector			Administrator
Performance Measure:	Attendance	Collusion	Inspections	Collusion	Inspections	Doctor Attendance
Big 5 Index		– –			++	++
Agreeableness		– –			+	
Conscientiousness	++	– –			+	+
Extroversion		– –		+		
Emotional Stability		– –			++	+
Openness						
PSM Index	++	– –		– –		
Attraction		– –			++	
Civic Duty	++	– –				
Commitment		– –		– –		
Compassion	+	– –				
Self Sacrifice	++	– –		–		
Social Justice		– –		–		

Notes: This table provides a summary of the results available in Figures 2.4 and 2.7 and Tables 2.3 and 2.4. + (++) indicates a positive relationship significant at the 10 percent (5 percent) level and – (– –) indicates a negative relationship significant at the 10 percent (5 percent) level.

formance. In addition, our data also allow us to examine the pattern of relationships between specific traits and performance.

Five of the seven tests linking conscientiousness—the personality trait of being responsible, hardworking, and desiring to do tasks well—to performance are statistically significant with at least 90 percent confidence and always in the predicted direction. This trait appears to be important for doctors, inspectors, and senior officials alike. Similarly, four of the tests linking emotional stability—the personality trait of being calm and not reactive to stress—are statistically significant with at least 90 percent confidence and also always in the right direction. This trait also matters for doctors, inspectors, and senior officials. It also is important for both the response of inspectors to a monitoring treatment which sharpened their incentives to perform inspections and for the response of senior officials to information about the absence of their subordinates. These interventions likely resulted in a greater volume of work for these officials. It appears that those who score better on a trait meant to capture the ability to respond to new pressures indeed do better when work demands are increased. By contrast, openness—the trait associated with curiosity, a desire for adventure, and valuing a variety of experience—does not predict performance in any of the seven tests performed. This may be because, in a structured bureaucracy, where a person scores on this trait does not matter for performance.

Our goal in this investigation is to examine whether differences between public sector workers affect the quality of services. We find a relationship between the non-cognitive attributes of workers and their performance on objective service delivery tasks for 30 of the 77 tests we perform. We interpret this as supporting the view that, in addition to improving incentives, we can also improve government performance by improving the non-cognitive qualities of individuals working in public service.

2.6 Acknowledgements

Chapter 2, in part, is currently being prepared for submission for publication of the material. Callen, Michael; Gulzar, Saad; Khan, Yasir; Hasanain, Ali; Rezaee, Arman. Support was provided by the International Growth Centre (IGC) state capabilities program, the IGC Pakistan country office, and the University of California Office of the President Lab Fees Research Program Grant #235855. Callen was supported by grant #FA9550-09-1-0314 from the Air Force Office of Scientific Research.

2.7 Chapter 2 Appendix

2.7.1 Appendix Tables

Table 2.6: Doctor and Health Inspector Personality Summary Statistics (Control Districts)

	Mean	SD	P10	P50	P90	Obs
PANEL A: Doctor Personality Summary Statistics						
Big Five Personality Traits						
Big Five Index	0.04	0.79	-0.99	0.05	1.14	192
Agreeableness	3.57	0.66	2.67	3.67	4.42	192
Conscientiousness	4.02	0.55	3.33	4	4.75	192
Extroversion	3.69	0.48	3.17	3.67	4.33	192
Emotional Stability	-2.54	0.70	-3.50	-2.50	-1.67	192
Openness	2.92	0.44	2.42	2.92	3.50	192
Public Service Motivation						
PSM Index	0.02	0.67	-0.83	-0.01	0.92	192
Attraction	3.46	0.60	2.60	3.40	4.20	192
Civic Duty	4.22	0.53	3.43	4.29	5	192
Commitment	3.79	0.45	3.29	3.86	4.29	192
Compassion	3.55	0.53	2.88	3.50	4.25	192
Self Sacrifice	4.09	0.60	3.38	4.12	4.88	192
Social Justice	3.96	0.59	3.20	4	4.60	192
Performance						
Present (=1)	0.43	0.50	0	0	1	637
PANEL B: Inspector Personality Summary Statistics						
Big Five Personality Traits						
Big Five Index	0.02	0.75	-1.25	0.11	1.04	48
Agreeableness	3.66	0.54	2.67	3.79	4.25	48
Conscientiousness	4.12	0.54	3.33	4.21	4.75	48
Extroversion	3.73	0.46	3.17	3.70	4.33	48
Emotional Stability	-2.34	0.62	-3.25	-2.25	-1.58	48
Openness	3.11	0.35	2.67	3.17	3.58	48
Public Service Motivation						
PSM Index	0.07	0.61	-0.77	0.13	0.69	49
Attraction	3.57	0.57	2.80	3.60	4.25	49
Civic Duty	4.44	0.42	3.86	4.57	5	49
Commitment	3.97	0.37	3.43	3.86	4.50	49
Compassion	3.66	0.49	3	3.62	4.25	49
Self Sacrifice	4.40	0.45	3.86	4.50	5	49
Social Justice	4.20	0.43	3.60	4.20	5	49
Performance						
Inspected in the Last Two Months (=1)	0.56	0.50	0	1	1	557
PANEL C: Collusion						
Predicted Collusion (=1)	0.13	0.33	0	0	1	334

Notes: Sample for Panel A: doctors in control districts that completed the personalities survey module, given in waves 2 and 3 and during a special follow-up round. Sample for Panel B: health inspectors in control districts that completed the personalities survey module. Doctors and inspectors were only asked to complete the module once. Performance and collusion samples are clinic-wave observations in control districts across waves 1 through 3, where doctors are posted. Collusion is a dummy variable coded as 1 when a doctor is reported absent in both survey waves 2 and 3 but is reported as present by health inspectors during every visit between the launch of the program and present (up to 50 visits).

Table 2.7: Senior Health Official Personality Summary Statistics (Control Districts)

	Mean	SD	P10	P50	P90	Obs
Big Five Personality Traits						
Big Five Index	0.07	0.74	-0.89	0.47	0.72	16
Agreeableness	3.75	0.59	3.17	3.88	4.33	16
Conscientiousness	4.10	0.51	3.42	4.25	4.67	16
Extroversion	3.80	0.34	3.42	3.83	4.25	16
Emotional Stability	-2.34	0.53	-3.17	-2.09	-1.75	16
Openness	3.07	0.36	2.73	2.88	3.58	16
Public Sector Motivation						
PSM Index	0.20	0.63	-0.64	0.06	1.00	16
Attraction	3.73	0.61	3.00	3.50	4.80	16
Civic Duty	4.54	0.39	3.86	4.57	5.00	16
Commitment	3.95	0.35	3.57	4.00	4.43	16
Compassion	3.80	0.45	3.25	3.62	4.50	16
Self Sacrifice	4.51	0.34	4.00	4.56	4.88	16
Social Justice	4.16	0.42	3.60	4.10	4.80	16

Notes: Sample: senior health officials in control districts that completed the personalities survey module, given during a single round after the final wave of clinic visits. The Big Five and PSM traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively.

Table 2.8: Doctor Personality and Doctor Attendance

	Doctor Present (=1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: Big Five Personality Traits							
Big Five Index		0.037					
		(0.034)					
Agreeableness			0.006				
			(0.023)				
Conscientiousness				0.055**			
				(0.026)			
Extroversion					0.045*		
					(0.025)		
Emotional Stability						0.025	
						(0.024)	
Openness							-0.017
							(0.023)
Mean of Dependent Variable		0.493	0.493	0.493	0.493	0.493	0.493
# Observations		479	479	479	479	479	479
# Clinics		190	190	190	190	190	190
R-Squared		0.192	0.190	0.197	0.195	0.191	0.190
PANEL B: Public Service Motivation							
PSM index	0.074**						
	(0.036)						
Attraction		0.029					
		(0.025)					
Civic Duty			0.067**				
			(0.030)				
Commitment				0.030			
				(0.026)			
Compassion					0.008		
					(0.027)		
Self Sacrifice						0.052**	
						(0.025)	
Social Justice							0.027
							(0.022)
Mean of Dependent Variable	0.493	0.493	0.493	0.493	0.493	0.493	0.493
# Observations	479	479	479	479	479	479	479
# Clinics	190	190	190	190	190	190	190
R-Squared	0.196	0.192	0.199	0.192	0.190	0.197	0.192

Notes: $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the clinic level reported in parentheses. All regressions include Tehsil (sub-district) and survey wave fixed effects. Sample: control district clinics for which doctor personality data is available and a doctor is posted. The Big Five and PSM traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). All personality traits are then normalized across doctors. The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively.

Table 2.9: Doctor Personality and Estimated Doctor-inspector Collusion

	Doctor-inspector Collusion (=1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: Big Five Personality Traits							
Big Five Index		-0.098***					
		(0.031)					
Agreeableness			-0.083***				
			(0.026)				
Conscientiousness				-0.058***			
				(0.021)			
Extroversion					-0.061***		
					(0.022)		
Emotional Stability						-0.063***	
						(0.021)	
Openness							-0.012
							(0.025)
Mean of Dependent Variable		0.103	0.103	0.103	0.103	0.103	0.103
# Observations		273	273	273	273	273	273
# Clinics		273	273	273	273	273	273
R-Squared		0.389	0.399	0.373	0.377	0.378	0.347
PANEL B: Public Service Motivation							
PSM index	-0.123***						
	(0.036)						
Attraction		-0.054**					
		(0.022)					
Civic Duty			-0.051**				
			(0.022)				
Commitment				-0.069***			
				(0.024)			
Compassion					-0.066***		
					(0.023)		
Self Sacrifice						-0.066***	
						(0.021)	
Social Justice							-0.049**
							(0.022)
Mean of Dependent Variable	0.103	0.103	0.103	0.103	0.103	0.103	0.103
# Observations	273	273	273	273	273	273	273
# Clinics	273	273	273	273	273	273	273
R-Squared	0.408	0.371	0.371	0.388	0.381	0.382	0.366

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard errors clustered at the clinic level reported in parentheses. All regressions include Tehsil (sub-district) and survey wave fixed effects. Sample: treatment district clinics for which doctor personality data is available and a doctor is posted. All personality traits are normalized. The Big Five and PSM traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). All personality traits are then normalized across doctors. The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively. Collusion is a dummy variable coded as 1 when a doctor is reported absent in both survey waves 2 and 3 but is reported as present by health inspectors during every visit between the launch of the program and present (up to 50 visits).

Table 2.10: Doctor Personality Measure Predictions Compared to Other Covariates

	(1)	(2)	(3)	(4)	(5)
	Doctor Present (=1)				
Distance to Hometown (KM)	-0.000*		-0.000*		-0.000*
	(0.000)		(0.000)		(0.000)
Tenure in Department of Health (Years)	0.000		-0.000		-0.000
	(0.000)		(0.000)		(0.000)
Tenure at Clinic (Years)	-0.001		-0.001		-0.001
	(0.000)		(0.001)		(0.001)
Big Five Index		0.037	0.036		
		(0.034)	(0.033)		
PSM Index				0.074**	0.075**
				(0.036)	(0.034)
Mean of Dependent Variable	0.502	0.493	0.484	0.493	0.484
# Observations	514	479	471	479	471
# Clinics	212	190	187	190	187
R-Squared	0.180	0.192	0.193	0.196	0.198
	Doctor-inspector Collusion (=1)				
Distance to Hometown (KM)	-0.000		-0.000		0.000
	(0.000)		(0.000)		(0.000)
Tenure in Department of Health (Years)	-0.000		0.000		0.000
	(0.000)		(0.000)		(0.000)
Tenure at clinic (Years)	-0.000		-0.001		-0.001
	(0.001)		(0.001)		(0.001)
Big Five Index		-0.097***	-0.099***		
		(0.031)	(0.031)		
PSM Index				-0.123***	-0.122***
				(0.036)	(0.037)
Mean of Dependent Variable	0.112	0.103	0.100	0.103	0.100
# Observations	295	273	269	273	269
# Clinics	295	273	269	273	269
R-Squared	0.333	0.388	0.413	0.408	0.428

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard errors clustered at the clinic level reported in parentheses. All regressions include Tehsil (sub-district) and survey wave fixed effects. Sample: Clinics for which doctor personality data is available and a doctor is posted. Panel A is restricted to control clinics, Panel B to treatment. The Big Five and PSM traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). All personality traits are then normalized across doctors. The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively. Collusion is a dummy variable coded as 1 when a doctor is reported absent in both survey waves 2 and 3 but is reported as present by health inspectors during every visit between the launch of the program and present (up to 50 visits).

Table 2.11: Doctor Attendance and Health Service Provision (Control Districts)

	Number of Outpatients Seen
	(1)
Present (=1)	201.250***
	(51.557)
Mean of Dependent Variable	1071.704
# Observations	783
# Clinics	419
R-Squared	0.419

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard errors clustered at the clinic level reported in parentheses. Regression includes Tehsil (sub-district) and survey wave fixed effects. Sample is limited to clinics in control districts which keep records of outpatient visits (419 of 425). The number of outpatients seen is in the total for each month prior to our independent visits. Present is a dummy variable equal to one if the clinic's doctor was present during the same independent visits.

Table 2.12: Health Inspector Personality and Inspections

	Health Inspection in Last Two Months (=1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: Big Five Personality Traits							
Big Five Index		-0.062					
		(0.044)					
Agreeableness			-0.024				
			(0.030)				
Conscientiousness				-0.043*			
				(0.024)			
Extroversion					-0.043		
					(0.034)		
Emotional Stability						-0.060*	
						(0.032)	
Openness							0.004
							(0.032)
Mean of Dependent Variable		0.589	0.589	0.589	0.589	0.589	0.589
# Observations		453	453	453	453	453	453
# Tehsils		45	45	45	45	45	45
R-Squared		0.164	0.161	0.164	0.164	0.166	0.160
PANEL B: Public Service Motivation							
PSM Index	-0.044						
	(0.052)						
Attraction		-0.004					
		(0.034)					
Civic Duty			-0.010				
			(0.023)				
Commitment				-0.024			
				(0.035)			
Compassion					-0.058		
					(0.057)		
Self Sacrifice						-0.018	
						(0.022)	
Social Justice							-0.014
							(0.035)
Mean of Dependent Variable	0.573	0.573	0.573	0.573	0.573	0.573	0.573
# Clinics	46	46	46	46	46	46	46
# Tehsils	466	466	466	466	466	466	466
R-Squared	0.191	0.189	0.189	0.190	0.193	0.190	0.189

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the health inspector level reported in parentheses. All regressions include district and survey wave fixed effects. Sample: control district clinics for which health inspector personality data is available and a doctor is posted. The Big Five and PSM traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). All personality traits are then normalized across inspectors. The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively.

Table 2.13: Health Inspector Personality and Estimated Doctor-inspector Collusion

	Doctor-inspector Collusion (=1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: Big Five Personality Traits							
Big Five Index		0.051					
		(0.044)					
Agreeableness			-0.001				
			(0.027)				
Conscientiousness				0.009			
				(0.022)			
Extroversion					0.055*		
					(0.028)		
Emotional Stability						0.007	
						(0.016)	
Openness							0.004
							(0.019)
Mean of Dependent Variable		0.092	0.092	0.092	0.092	0.092	0.092
# Observations		292	292	292	292	292	292
# Tehsils		48	48	48	48	48	48
R-Squared		0.148	0.144	0.144	0.159	0.144	0.144
PANEL B: Public Service Motivation							
PSM Index	-0.071**						
	(0.029)						
Attraction		-0.042					
		(0.030)					
Civic Duty			0.009				
			(0.018)				
Commitment				-0.050**			
				(0.019)			
Compassion					-0.020		
					(0.019)		
Self Sacrifice						-0.031*	
						(0.016)	
Social Justice							-0.035*
							(0.019)
Mean of Dependent Variable	0.095	0.095	0.095	0.095	0.095	0.095	0.095
# Observations	294	294	294	294	294	294	294
# Tehsils	49	49	49	49	49	49	49
R-Squared	0.160	0.153	0.149	0.167	0.151	0.155	0.157

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the health inspector level reported in parentheses. All regressions include district and survey wave fixed effects. Sample: control district clinics for which health inspector personality data is available and a doctor is posted. The Big Five and PSM traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). All personality traits are then normalized across inspectors. The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively. Collusion is a dummy variable coded as 1 when a doctor is reported absent in both survey waves 2 and 3 but is reported as present by health inspectors during every visit between the launch of the program and present (up to 50 visits).

Table 2.14: Health Inspector Personality and Inspections—Full Sample

	Health Inspection in Last Two Months (=1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: Big Five Personality Traits							
Big 5 index		-0.020					
		(0.028)					
Agreeableness			0.010				
			(0.020)				
Conscientiousness				-0.017			
				(0.017)			
Extroversion					-0.034		
					(0.025)		
Emotional stability						-0.041	
						(0.032)	
Openness							0.038
							(0.026)
Mean of dependent variable		0.635	0.635	0.635	0.635	0.635	0.635
# Observations		860	860	860	860	860	860
# Tehsils		49	49	49	49	49	49
R-Squared		0.180	0.180	0.180	0.182	0.182	0.182
PANEL B: Public Service Motivation							
PSM index	-0.000						
	(0.041)						
Attraction		-0.005					
		(0.027)					
Civic duty			0.013				
			(0.020)				
Commitment				0.018			
				(0.025)			
Compassion					-0.027		
					(0.025)		
Self Sacrifice						0.008	
						(0.017)	
Social justice							-0.022
							(0.024)
Mean of dependent variable	0.619	0.619	0.619	0.619	0.619	0.619	0.619
# Observations	885	885	885	885	885	885	885
# Tehsils	50	50	50	50	50	50	50
R-Squared	0.206	0.206	0.207	0.207	0.208	0.207	0.207

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the health inspector level reported in parentheses. All regressions include district and survey wave fixed effects. Sample: control district clinics for which health inspector personality data is available, regardless of whether or not a doctor is posted. The Big Five and PSM traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). All personality traits are then normalized across inspectors. The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively.

Table 2.15: Health Inspector Personality and Estimated Doctor-inspector Collusion—Full Sample

	Doctor-inspector Collusion (=1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: Big Five Personality Traits							
Big 5 index		0.102*					
		(0.051)					
Agreeableness			0.047				
			(0.032)				
Conscientiousness				0.051*			
				(0.026)			
Extroversion					0.040		
					(0.037)		
Emotional stability						0.020	
						(0.022)	
Openness							0.002
							(0.026)
Mean of dependent variable		0.194	0.194	0.194	0.194	0.194	0.194
# Observations		360	360	360	360	360	360
# Tehsils		51	51	51	51	51	51
R-Squared		0.183	0.179	0.181	0.178	0.175	0.173
PANEL B: Public Service Motivation							
PSM index	-0.016						
	(0.046)						
Attraction		-0.026					
		(0.033)					
Civic duty			0.040				
			(0.025)				
Commitment				-0.046**			
				(0.019)			
Compassion					-0.005		
					(0.023)		
Self Sacrifice						-0.005	
						(0.030)	
Social justice							0.002
							(0.025)
Mean of dependent variable	0.196	0.196	0.196	0.196	0.196	0.196	0.196
# Observations	362	362	362	362	362	362	362
# Tehsils	52	52	52	52	52	52	52
R-Squared	0.174	0.174	0.178	0.182	0.173	0.173	0.173

Notes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors clustered at the health inspector level reported in parentheses. All regressions include district and survey wave fixed effects. Sample: control district clinics for which health inspector personality data is available, regardless of whether a doctor is posted. Collusion is a dummy variable coded as 1 when a doctor is reported absent in both survey waves 2 and 3 but is reported as present by health inspectors during every visit between the launch of the program and present (up to 50 visits).

Table 2.16: Personalities and Health Inspections—Experimental Evidence, Unconditional on Doctor Being Posted

	Health Inspection in Last Two Months (=1)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: Big Five Personality Traits									
Monitoring (=1)		0.267**	0.141	0.126	0.165	0.134	0.144	0.105	0.143
		(0.129)	(0.119)	(0.107)	(0.103)	(0.105)	(0.117)	(0.111)	(0.115)
Monitoring x Big Five Index				0.234					
				(0.144)					
Monitoring x Agreeableness					0.104				
					(0.091)				
Monitoring x Conscientiousness						0.134			
						(0.100)			
Monitoring x Extroversion							0.042		
							(0.080)		
Monitoring x Emotional Stability								0.142	
								(0.087)	
Monitoring x Openness									0.165*
									(0.096)
Mean of Dependent Variable		0.651	0.673	0.673	0.673	0.673	0.673	0.673	0.673
# Observations		2173	1808	1808	1808	1808	1808	1808	1808
# Clinics		35	35	35	35	35	35	35	35
R-Squared		0.049	0.044	0.061	0.069	0.060	0.046	0.056	0.055
PANEL B: Public Service Motivation									
Monitoring (=1)	0.267**	0.152	0.141	0.147	0.143	0.137	0.150	0.158	0.138
	(0.129)	(0.116)	(0.112)	(0.108)	(0.116)	(0.105)	(0.114)	(0.124)	(0.111)
Monitoring x PSM Index			0.155						
			(0.153)						
Monitoring x Attraction				0.199**					
				(0.074)					
Monitoring x Civic Duty					-0.048				
					(0.070)				
Monitoring x Commitment						0.032			
						(0.078)			
Monitoring x Compassion							0.101		
							(0.093)		
Monitoring x Self Sacrifice								-0.034	
								(0.095)	
Monitoring x Social Justice									0.083
									(0.098)
Mean of Dependent Variable	0.651	0.664	0.664	0.664	0.664	0.664	0.664	0.664	0.664
# Observations	2173	1839	1839	1839	1839	1839	1839	1839	1839
# Clinics	35	35	35	35	35	35	35	35	35
R-Squared	0.049	0.045	0.053	0.066	0.046	0.057	0.049	0.046	0.052

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the district level reported in parentheses. All regressions include Tehsil (sub-district) and survey wave fixed effects and are not conditional on a doctor being posted. Column (1) reports average treatment effects on treatment and control district clinics. Columns (2) - (10) are limited to clinics in tehsils for which health inspector personality data is available. The Big Five and PSM traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). All personality traits are then normalized across inspectors. The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively.

Table 2.17: Inspector Personality Measure Experimental Results Compared to Other Covariates

	(1)	(2)	(3)	(4)	(5)	(6)
	Inspected in the Last Two Months (=1)					
Monitoring (=1)	0.178	1.015	-0.006	0.244	0.023	0.659
	(0.154)	(1.121)	(0.114)	(1.092)	(0.120)	(1.094)
Monitoring x Age (Years)		0.001		0.011		0.012
		(0.032)		(0.031)		(0.032)
Monitoring x Has Completed Higher Education (=1)		0.205		0.358*		0.296
		(0.147)		(0.148)		(0.155)
Monitoring x Tenure in Department of Health (Years)		-0.034		-0.027		-0.044
		(0.032)		(0.033)		(0.032)
Monitoring x Tenure as Inspector (Years)		0.028		0.019		0.023
		(0.024)		(0.024)		(0.029)
Monitoring x Distance to Hometown (KM)		0.047		0.085		0.086
		(0.027)		(0.050)		(0.049)
Monitoring x Inspector Reports Liking Current Ppost (=1)		-0.061		-0.058		-0.062
		(0.048)		(0.048)		(0.048)
Monitoring x PSM Index			0.351*	0.277		
			(0.133)	(0.167)		
Monitoring x Big Five Index					0.202	0.120
					(0.140)	(0.159)
Mean of Dependent Variable	0.642	0.645	0.656	0.504	0.649	0.503
# Observations	1331	1177	1145	1132	1164	1151
# Tehsils	35	33	34	33	34	33
R-Squared	0.048	0.095	0.069	0.103	0.057	0.098

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the district level reported in parentheses. All regressions include district and survey wave fixed effects. Sample: clinics for which health inspector personality data is available and a doctor is posted. The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively.

Table 2.18: Differential Clinic Flagging Effects by Senior Health Official Personality, Robustness to Cutoff

	Doctor Present (=1)					
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Big Five I ndex						
Clinic Flagged as Underperforming on Dashboard	0.100	0.094	0.099	0.086	0.146	0.159
	(0.067)	(0.067)	(0.073)	(0.072)	(0.103)	(0.098)
Flagged x Big Five Index		0.118		0.249*		0.402**
		(0.131)		(0.143)		(0.200)
Mean of Dependent Variable	0.521	0.521	0.528	0.528	0.480	0.480
# Observations	326	326	233	233	123	123
# Clinics	228	228	180	180	106	106
R-Squared	0.114	0.117	0.140	0.152	0.204	0.231
PANEL B: PSM Index						
Clinic Flagged as Underperforming on Dashboard	0.100	0.098	0.099	0.111	0.146	0.165
	(0.067)	(0.070)	(0.073)	(0.075)	(0.103)	(0.105)
Flagged x PSM Index		-0.016		0.082		0.124
		(0.108)		(0.117)		(0.169)
Mean of Dependent Variable	0.521	0.521	0.528	0.528	0.480	0.480
# Observations	326	326	233	233	123	123
# Clinics	228	228	180	180	106	106
R-Squared	0.114	0.114	0.140	0.142	0.204	0.208
Sample	Full	Full	Partial	Partial	Disc.	Disc.

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard errors clustered at the clinic level reported in parentheses. All regressions include district and survey wave fixed effects and condition on a doctor being posted. Clinics were flagged as underperforming if three or more of the seven staff were absent in one or more health inspections of the clinic fifteen to fourty-five days prior to an unanounced visit by our survey enumerators. The sample is limited to Monitoring the Monitor treatment districts due to the necessity of the web dashboard for flagging clinics. In addition, columns (3) and (4) restrict the sample to those clinics where only four or less staff were absent. We call this sample the "partial" sample. Columns (5) and (6) restrict the sample to those clinics where only two or three staff were absent. We call this sample the "discontinuity" sample. The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively.

Table 2.19: Differential Clinic Flagging Effects by Senior Health Official Personality, Semi-parametric

	(1)	(2)
	Inspected in the Last Two Months (=1)	
Clinic Flagged as Underperforming on Dashboard	-0.143	0.074
	(0.193)	(0.170)
Flagged x Big Five Index Second Quartile (=1)	0.250	
	(0.251)	
Flagged x Big Five Index Third Quartile (=1)	0.396	
	(0.264)	
Flagged x Big Five Index Fourth Quartile (=1)	0.650**	
	(0.278)	
Flagged x PSM Index Second Quartile (=1)		0.497**
		(0.237)
Flagged x PSM Index Third Quartile (=1)		-0.068
		(0.239)
Flagged x PSM Index Fourth Quartile (=1)		0.308
		(0.261)
Mean of Dependent Variable	0.520	0.520
# Observations	123	123
# Clinics	106	106
R-Squared	0.244	0.225

Notes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors clustered at the clinic level reported in parentheses. All regressions include district and survey wave fixed effects and condition on a doctor being posted. Clinics were flagged as underperforming if three or more of the seven staff were absent in one or more health inspections of the clinic fifteen to fourty-five days prior to an unanounced visit by our survey enumerators. The sample is limited to Monitoring the Monitor treatment districts due to the necessity of the web dashboard for flagging clinics. In addition, all columns restrict the sample to those clinics where only two or three staff were absent.

Table 2.20: Differential Senior Health Official Time Use by Personality

	Share of Time Senior Health Official Spent Monitoring Health Facilities						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Number of Clinics Flagged as Underperforming on Dashboard	0.009	0.014***	0.011**	0.012**	0.010*	0.012*	0.008
	(0.006)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
# Flagged x Big Five Index		0.031*					
		(0.016)					
# Flagged x Agreeableness			-0.000				
			(0.007)				
# Flagged x Conscientiousness				0.015*			
				(0.008)			
# Flagged x Extroversion					0.005		
					(0.007)		
# Flagged x Emotional Stability						0.011	
						(0.008)	
# Flagged x Openness							0.011
							(0.007)
Mean of the Dependent Variable	0.097	0.097	0.097	0.097	0.097	0.097	0.097
# Observations	17	17	17	17	17	17	17
R-Squared	0.124	0.361	0.160	0.413	0.156	0.188	0.289

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors reported in parentheses. Sample limited to senior health officials in treatment districts. Clinics were flagged as underperforming if three or more of the seven staff were absent in one or more health inspections of the clinic fifteen to fourty-five days prior to an unanounced visit by our survey enumerators. The sample is limited to Monitoring the Monitor treatment districts due to the necessity of the web dashboard for flagging clinics. The number flagged is the total number of clinics flagged in each district prior to our second endline (when we also collected senior health official personality and time use). Each regression also contains a control for the personality measure uninteracted. The Big Five traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). All personality traits are then normalized across senior health officials. The Big Five Index is a z-score averages of the five Big Five traits.

Table 2.21: Differential Clinic Flagging Effects by Senior Health Official Personality Compared to Other Covariates

	(1)	(2)	(3)	(4)	(5)	(6)
	Doctor Present (=1)					
Clinic Flagged as Underperforming on Dashboard	0.146	-1.528	0.159	0.800	0.165	1.917
	(0.103)	(2.640)	(0.098)	(2.564)	(0.105)	(3.613)
Flagged x Age (Years)		0.058		0.028		0.038
		(0.055)		(0.059)		(0.061)
Flagged x Has Completed Higher Education (=1)		0.326		0.241		0.215
		(0.290)		(0.248)		(0.314)
Flagged x Tenure in Department of Health (Years)		-0.058		-0.080		-0.120
		(0.084)		(0.079)		(0.072)
Flagged x Tenure as Official (Years)		-0.014		0.030		0.031
		(0.039)		(0.041)		(0.047)
Flagged x Distance to Hometown (KM)		0.011		-0.048		-0.039
		(0.030)		(0.034)		(0.037)
Flagged x Official Reports Liking Current Post (=1)		0.008		-0.002		-0.068
		(0.048)		(0.045)		(0.071)
Flagged x PSM Index			0.402*	0.552*		
			(0.200)	(0.242)		
Flagged x Big Five Index					0.124	0.452
					(0.169)	(0.347)
Mean of Dependent Variable	0.520	0.520	0.520	0.520	0.520	0.520
# Observations	123	123	123	123	123	123
# Clinics	106	106	106	106	106	106
R-Squared	0.204	0.225	0.231	0.245	0.208	0.235

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the clinic level reported in parentheses. All regressions include district and survey wave fixed effects and condition on a doctor being posted. Clinics were flagged as underperforming if three or more of the seven staff were absent in one or more health inspections of the clinic fifteen to fourty-five days prior to an unanounced visit by our survey enumerators. The sample is limited to Monitoring the Monitor treatment districts due to the necessity of the web dashboard for flagging clinics. In addition, the sample is restructed to those clinics where only two or three staff were absent. We call this sample the "discontinuity" sample. The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively.

Table 2.22: Results Summary

Alternative Hypothesis:	Personality Predicts Performance				Personality Predicts Monitoring Treatment Heterogeneity	Personality Predicts Information Treatment Heterogeneity
Public Actor:	Doctor		Inspector			Administrator
Performance Measure:	Attendance	Collusion	Inspections	Collusion	Inspections	Doctor Attendance
Big 5 Index		- -			++	+
Agreeableness		-				
Conscientiousness	+	-				
Extroversion	+					
Emotional Stability					+	
Openness						
PSM Index	++	- -		- -		
Attraction					+	
Civic Duty	+	-				
Commitment		-		-		
Compassion		-				
Self Sacrifice	+	-				
Social Justice						

Notes: +/- indicates sign of coefficients from regression reported in Figures 2.4 and 2.7 and Tables 2.3 and 2.4. For information on each specific hypothesis test, see these. The number of + or -'s representats the p-value from the reported hypothesis test (one corresponds to a p-value < 0.1, two to a p-value < 0.05). All p-values are corrected for multiple hypothesis testing using the False Discovery Rate procedure outlined in Benjamini and Hochberg (1995). P-values for the 14 index hypotheses are corrected separately than the p-values for the 55 trait hypotheses.

2.7.2 Appendix Figures

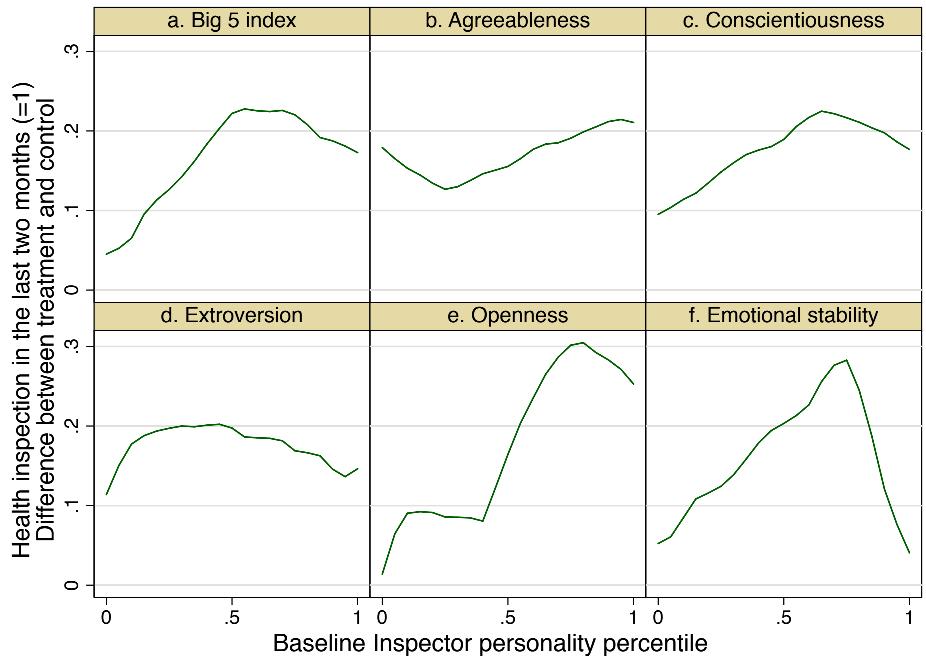

Figure 2.11: Health Inspector Non-parametric Heterogeneous Effects, Trait-by-Trait, Big Five

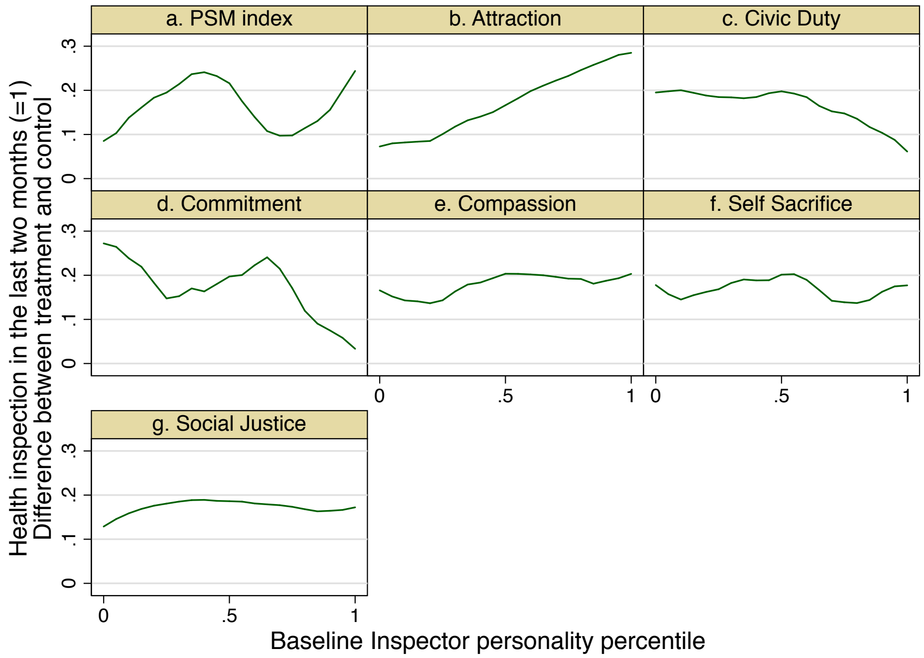

Figure 2.12: Health Inspector Non-parametric Heterogeneous Effects, Trait-by-Trait, PSM

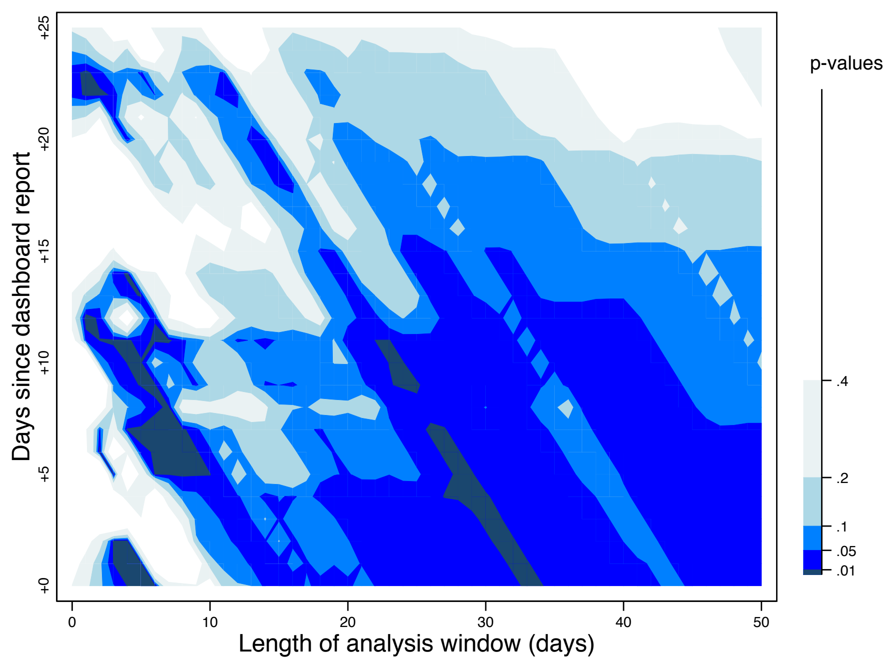

Figure 2.13: Robustness to Different Windows for Flagging—Big Five Index

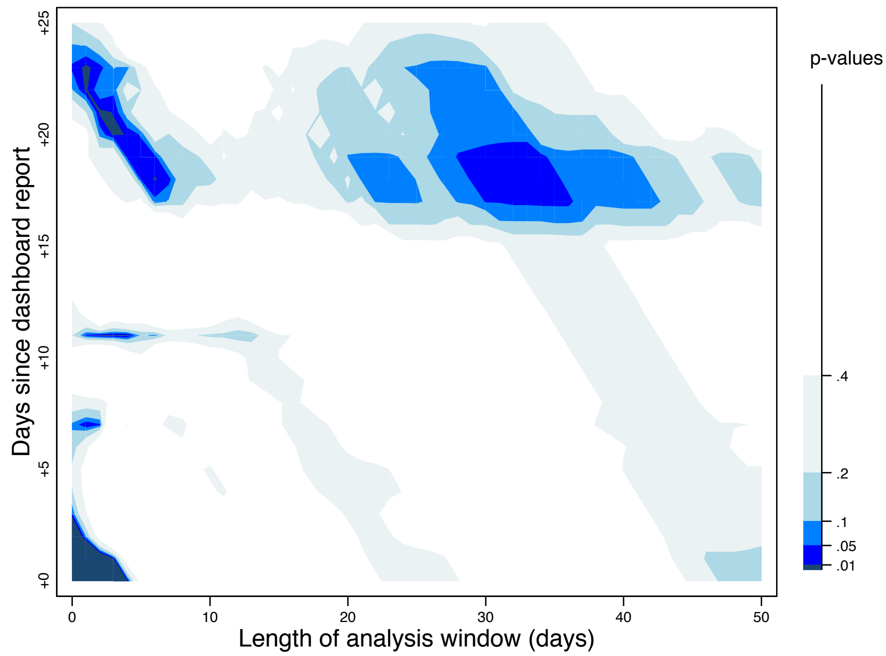

Figure 2.14: Robustness to Different Windows for Flagging—PSM Index

2.7.3 Model Extension

As with the model in Section 2.2, let our personality measures represent a worker's type, θ, with cumulative distribution $F(\theta)$. Let performance be the binary decision that a doctor or health inspector makes of whether to show up to work or to shirk. If a worker chooses to work, he receives a fixed salary of W and incurs a cost of effort of $\lambda(\theta)$. If a worker chooses to shirk, he exerts no effort and receives the fixed salary with probability $1 - p$ and an arbitrarily small punishment c with probability p as well as an outside option of Q. However, let us now assume that the outside option is a function of θ. Thus we have the following updated indifference condition:

$$W - \lambda(\theta) = (1 - p)W - pc + Q(\theta) \tag{2.7}$$

Though it is still straightforward to see here that an increase in p weakly increases the probability that a given worker will choose to work, it is not as straightforward, to determine either the status quo correlation between θ and performance or which types from the distribution of θ will respond to a given increase in p. To get traction on this, we will make two analagous assumptions. Assume that $\frac{\partial\lambda(\theta)}{\partial\theta} > 0, asbefore$, and that $\frac{\partial Q(\theta)}{\partial\theta} > 0$.

Given these assumptions, we can plot the net payoff to working versus the net payoff to shirking before and after an increase in p under various scenarios. Both Figure 2.15 and 2.16 show a case when the $\lambda(\theta)$ function is linear and the $Q(\theta)$ function is convex in θ.[53]

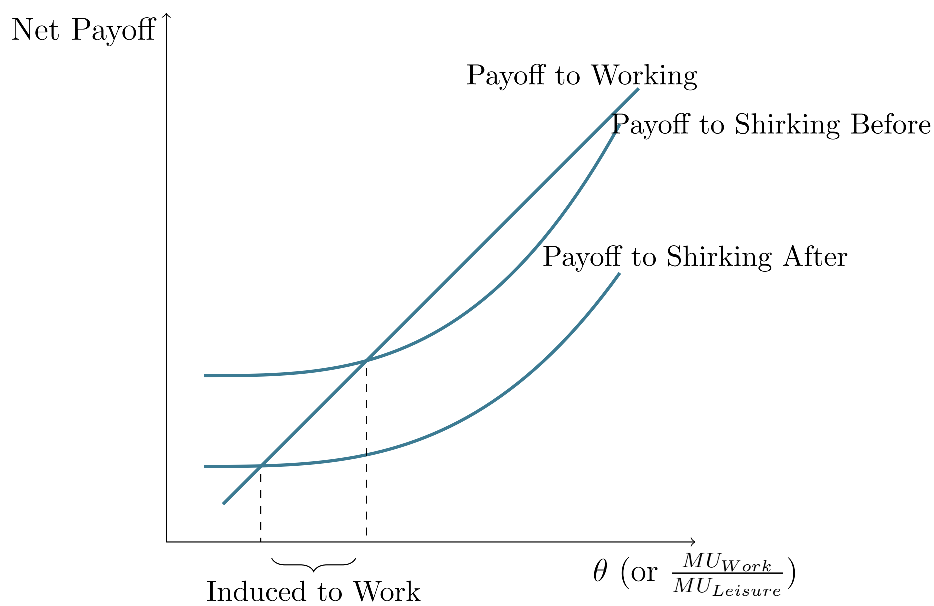

Figure 2.15: Effect of an Increase in Detection Probability on the Decision to Work or Shirk

These figures allow us to make several important points. First, we can see that in both figures an increase in incentives to work induces a range of workers in the

[53]Note that the case when both functions are linear is very unlikely to be accurate, while the case when both functions are strictly convex, while likely more accurate, does not lead to any additional intuition (both presented cases would hold so long as the $\lambda(\theta)$ function has less curvature than the $Q(\theta)$ function over the relevant range).

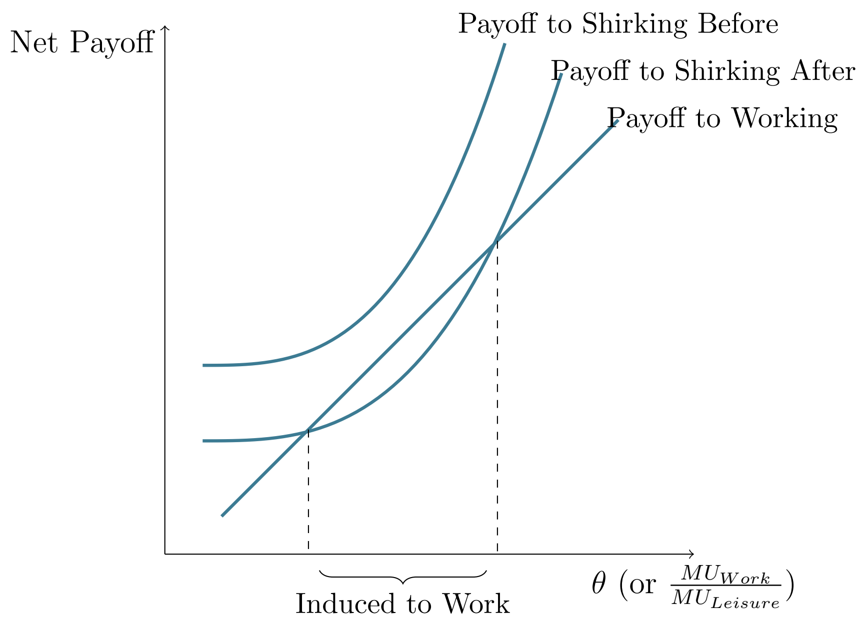

Figure 2.16: Effect of an Increase in Detection Probability on the Decision to Work or Shirk

middle of the personality type distribution to work. Second, we can see that in the second figure, before an increase in p no one chooses to work. This highlights that the existence of a relationship between performance and personality type is subject to the outside option for some personality types being sufficiently low. More generally, the difference between the two figures highlights the ambiguity in correlation between performance and personality type under a fixed p. In the first figure, all workers above a certain marginal worker will choose to work, with the marginal worker shifting to the left after p is increased. This would create a positive correlation between personality type and working under the status quo and a positive correlation between personality type and responding to an increase in p by switching from shirking to working. Where as in the second figure, the gains to the outside option for the highest personality types overcome the gains to those types for working even after p is increased sufficient to induce some personality types to work, causing the best personality types to join the worst personality types in shirking. This would lead to an ambiguous correlation between personality type and working.

2.7.4 The Effects of Improved Doctor Personality on Health Service Provision

To understand the impact of personality on health service provision, we will walk through a hypothetical exercise—imagine that through changes in hiring practices or through encouraging the improvement of personality traits, we were able to replace the bottom 25% of doctors in terms of their Big Five index with an average doctor in each respective measure. How would this impact health service provision in Punjab?

On average, the bottom 25% of doctors in our sample in control groups have a normalized Big Five index of -0.95 standard deviations. The mean doctor in the control groups has a normalized Big Five index of 0.05 standard deviations. Thus replacing the average bottom quartile doctor with an average doctor would raise doctor Big Five index by 1 standard deviation. This is then associated with 3.7 percentage point increase in doctor attendance respectively (combining the average Big Five index increase with the coefficient from Appendix Table 2.8). Now combining with our our results from Appendix Table 2.11, a 3.7 percentage point increase in doctor attendance is associated with an average increase of 7.45 outpatients seen per month. There are approximately 2496 total clinics in Punjab. Thus if we were to replace the bottom 25% of doctors in our sample with average doctors in terms of Big Five index, we would expect to increase total outpatients seen per month in Punjab by 4646.46 visits per month. This is a very meaningful number. And this is only half of what we would expect if we did the same exercise in terms of conscientiousness, the personality trait with the strongest correlational result for doctors. In summary:

Table 2.23: Doctor Attendance and Health Service Provision Appendix Exercise

	Personality measure	
	Big Five Index	Conscientiousness
Mean Score of Average Doctor in Bottom Quartile	-0.95 SDs	-1.31 SDs
Mean Score of Average Doctor in Entire Sample	0.05 SDs	0.06 SDs
Average Increase in Score	1 SDs	1.37 SDs
Average Increase in Attendance per SD of Score	3.7 PPs	5.5 PPs
Average Increase in Attendance as a Result of Increase in Score	3.7 PPs	7.5 PPs
Total Outpatients Seen per Month When Doctor was Present	201.25	201.25
Average Increase in Outpatients as a Result of Increase in Score	7.45	15.17
Approximate Total Number of Clinics in Punjab	2496	2496
25% of Total Clinics	624	624
Estimated Increase in Outpatients Per Month from Increase in Score	4646.46	9468.73

Notes: Reported standard deviations, percentage points, and regression coefficients all come from results above.

Chapter 3

Choosing Ungoverned Space: Pakistan's Frontier Crimes Regulation

3.1 Introduction

Territory with little or no effective state presence—ungoverned space—persists in many developing countries. In addition to having few state services, these areas also provide room for terrorists, smugglers, drug manufacturers, and criminals to operate, creating negative externalities locally and globally. Pakistan has many such areas, and has for over a century, as both a British colony and an independent nation. This ungoverned space in North-Western Pakistan has been set forth in the Frontier Crimes Regulation (FCR) of 1901, a system under which governance was largely left under tribal control. This law cleanly delineates areas with and without state institutions, and allows for documenting how these areas have changed over time, providing an opportunity to study the determinants of state control. We study one key predictor of the extent of the FCR jurisdiction over time—potential agricultural revenue—thereby contributing to the understanding of how and when states absorb ungoverned tracts.

During colonial rule, the British divided Pakistan into two main regions. The first was the Raj—areas where the British built modern political and bureaucratic institutions. This included a modern legal system, a tax system, a civil service, and an army. The second was governed according to the Frontier Crimes Regulation (FCR).[1] Here the British put a small number of "political agents" in charge of large tribal areas with almost no colonial institutions backing them. Instead of the Raj system, institutions already in existence were given the force of law, and traditional local councils, or Jirgas, made most legal decisions. Under FCR the political agents could appoint Jirga members and collectively punish tribes for the behavior of their

[1] There were also a few 'princely states" which held parts of modern-day Pakistan, including most notably Kashmir, but they were not nearly so prominent as in areas that are now within India.

members, but there were few other institutions. As a result of the British division, independence and subsequent partition left roughly half of modern-day Pakistan effectively ungoverned by the state. Over time, all of Pakistan has been removed from the FCR except for the small regions along the Afghan border known as the Federally Administered Tribal Areas (FATA) and a few small Provincially Administered Tribal Areas (PATA).[2]

There have been many empirical attempts to understand the initial choice to govern a space during colonial times. Several competing hypotheses have been offered to explain the broad patterns in the historical record: (i) the availability of resources, and the ease with which they can be extracted, determine the initial set of institutions (Diamond, 1998; Gallup et al., 1999; McArthur and Sachs, 2001; Acemoglu et al., 2001); (ii) natural terrain, and the military advantage it affords indigenous groups, make full colonization impractical in some regions (Fearon and Laitin, 2003; Nunn and Puga, 2012); and (iii) it is both efficient, and easier, to maintain order in such regions through a system of indirect governance (Padro i Miquel and Yared, 2012; Scott, 2009).[3] All three perspectives are consistent with the British decision to set up minimal governance institutions in areas initially under the FCR.

Turning to how state presence changes over time within internationally-recognized borders, there are several additional hypothesis pertaining to why a state may maintain or roll-back ungoverned space. Acemoglu et al. (2013) put forth a model in which individuals and/or parties push to add or remove areas from the formal state based on a vote cost-benefit analysis. Similarly, a literature on constrained kleptocracies examines situations in which it is optimal for kleptocrats to not control their entire

[2]These areas have periodically provide safe haven to domestic and international terrorists. Training facilities operate openly and with impunity to this date in some of them.

[3]Note Scott (2009) points out that peripheries of countries in South East Asia are typically poorer than the core areas of the country. In Pakistan's FATA, however, households in ungoverned space have high incomes relative to the country average (Blair et al., 2013).

territory (Grossman and Noh, 1990, 1994). And even well-intentioned governments may chose not to contest rebel control when the expenditures required to efficiently do so are high relative to the costs rebels can impose (Berman et al., 2011) or when there are rents to gained from having limited ungoverned space within one's territory (Felter, 2006; Bapat, 2011).

The decision to control territory is complex. A government must consider the economic and geostrategic benefits, the terrain, distance from the capital, and other factors relating to the cost of control, as well as the consent of those living in the territory to be governed. A broad literature concerning the size, shape, and number of countries in international equilibrium provides useful guidance, though it was developed to analyze the distribution of nation states (see e.g. Friedman, 1977; Alesina and Spolaore, 1997; Alesina and Wacziarg, 1998; Alesina and Spolaore, 2003). Larger political jurisdictions reduce the per capita cost to providing non-rival public goods, make income taxes more appealing relative to customs taxes (Easterly and Rebelo, 1993), and, if international trade is imperfect, increase economic activity and opportunities for co-insurance through market integration. Importantly for our context, these models highlight heterogeneous policy preferences as a central cost to integration. Often, stable equilibria feature states shaped to optimally balance the benefits of integration against the costs of preference heterogeneity.

Our paper concerns a related question, the time path of integration of peripheral areas into the state and how the rate of integration changes in response to technological innovations. We test a simple explanation for ungoverned space. Innovations to production technology increase the benefits of integration. With costs unchanged, this has two implications. First, the technological change should increase the amount of territory governed by the central authority. Second, since costs are heterogeneous within states, areas that benefit relatively more should be more likely to be integrated. Our core objective in this paper is to empirically test this logic for the integration of ungoverned space.

Production innovations could lead to greater integration through several mechanisms. First, such innovations increase potential tax revenues, potentially increasing the state's interest in controlling the territory. Second, access to public goods provided by the central authority, such as roads, access to courts and greater protection of property rights, and access to government support for agricultural production—a highly relevant consideration during the state-led Green Revolution in Pakistan—may increase interest in integration. In the language of Alesina and Spolaore (1997), policy preferences in peripheral regions will shift to be more aligned with the central authority's policy, resulting in a new equilibrium with greater integration of peripheral regions.

Several features of the process of integration suggest the latter mechanism may be at work. Pakistan gradually integrated outlying regions, removing the FCR successfully over the course of more than a century. The process almost never resulted in violent conflict, and was broadly supported both by the state and by citizens living in peripheral regions. While far from conclusive, this suggests that innovation shifts, particularly the episode we study between 1962 and 1965, shifted the equilibrium in such a way that the government, as well as citizens and indigenous authorities in peripheral areas, preferred integration.

We test this cost-benefit logic in two stages. First, we make use of geo-spatial information and crop suitability data from the Food and Agriculture Organization of the United Nations to study why the British chose to apply FCR to over half of modern-day Pakistan in 1901. Once we condition for proxies for local productivity and the cost of imposing state institutions we find no correlation between a sub-district's crop suitability for wheat (the main crop influenced by the Green Revolution in

Pakistan) and the initial British choice to apply FCR. This first result is correlational. It would be consistent with our hypothesis that increased potential revenue should have increased the British's desire to govern many parts of Pakistan if it was the costs of implementing institutions that was much more binding initially. Preliminary analysis suggests this may be the case. While the unconditional correlation between crop suitability for wheat and FCR application is positive, the conditional correlation is zero once we control for proxies for the costs and benefits of exercising control. We are in the process of collecting additional data to directly measure productivity in the late-19th century to confirm this result holds.

Next, we exploit the differential impact of the Green Revolution by crop suitability to understand Pakistan's decisions to continue to apply or to roll FCR back across parts of the country throughout the 1960s and 1970s. The Green Revolution in South Asia is widely understood to have increased productivity for wheat more in marginal areas than in already-productive regions, it mitigated the importance of crop suitability and thus caused lower-suitability sub-districts to 'catch-up' to other districts in potential revenue extraction. Because the FCR's original application was conditionally-independent of crop-suitability, the Green Revolution created a plausibly exogenous differential increase in agricultural land value. Places that were marginally suitable for wheat saw their value increase more at a specific point in time than areas which were highly suitable for wheat.

Exploiting this variation we find that an increase in crop suitability from 'medium' to 'good' increased a sub-district's probability of being left ungoverned by almost 10 percentage points following the Green Revolution relative to before. Lower-suitability districts were more likely to switch from expected revenue negative to positive as a result of the Green Revolution, and these districts were relatively more likely to be integrated into the state . And our results suggest a large effect of land value on integration. A one unit increase in crop suitability is associated with a 9.6 percentage points differential increase in the likelihood that FCR continues to be applied to a sub-district following the Green Revolution. This result has a causal interpretation to the extent that: (i) the initial decision to apply the FCR was conditionally-independent of wheat crop suitability (which it appears to have been); and (ii) the timing of the Green Revolution varietals' introduction in Pakistan was exogenous to planned changes in the extent of the FCR.

These results are valuable because they provide microeconomic evidence on the importance of extractable land value, and technology-driven changes in land value, for the choice to integrate peripheral areas into the state and because they provide additional evidence on the importance of the Green Revolution in South Asia. But they are especially important because they provide evidence that technological change can lead to ungoverned spaces being folded into country's cores without civil war or serious violence. The parts of Pakistan that still have FCR today are, of course, the most resistant to government control, but so were many parts of the sub-districts that were brought into the government in the 1970s. Yet what was stopping the government from integrating them, and those living in these peripheral areas from consenting to integration without violence, at least in part, a simple cost-benefit calculation.

This paper proceeds as follows. Section 3.2 provides additional background on the FCR. Section 3.3 outlines our data, Section 3.4 describes our empirical strategy, Section 3.5 presents results, and Section 3.6 concludes.

3.2 Background

3.2.1 The Frontier Crimes Regulation, Through Independence (1901-1947)

In the 1840s, the British began to replace the Sikh government in Punjab with the same colonial institutions that were taking hold across the British Raj—tax collectors, police, a modern legal system, and other bureaucratic structures. However, they met limited success in what was to become the North Western Frontier Province (NWFP), in at least two important ways.[4] First, much of the area was operating at a deficit due to limited crop yields and heavy security expenses. Second, the British legal system, being codified throughout India at the time through the 1860 Indian Penal Code and the Code of Criminal Procedure, was vehemently resisted by local Pashtun clan leaders and other established elites in favor of a customary legal system. Among other major differences, this customary system forgave crimes for honor reasons, including killings. Such differences were highly publicized, especially in cases involving women.[5]

After multiple decades of struggle, the British eventually decided to stop fighting the customary legal system in favor of appropriating it in what would be codified in 1901 as the Frontier Crimes Regulation (FCR). This regulation put a single 'political agent', appointed by the local Governor, in charge of the entire region. Criminal cases were to be first sent to a local council of elders, or Jirga, for trial. The political agent would then approve of the Jirga's ruling or could overturn it. Convicted crim-

[4]Initially, these areas were the districts of Hazara, Peshawar, Kohat, Bannu, Dera Ismail and Dera Ghazi Khan in the Punjab province. These and several other districts were then made into the NWFP in November, 1901. NWFP was renamed Khyber-Pakhtunkhwa in 2010.

[5]Nichols (2013).

inals were not allowed appeals. And importantly, Jirgas could not sentence anyone to death. The Jirgas and the political agent could, however, pass collective judgment on communities, or punish relatives of those convicted, rulings that were very much customary and would not be allowed in the modern British legal system.

Perhaps of equal importance, with this unique legal system in the NWFP came a profound lack of other institutions. Tax collection was minimal (the political agent was also in charge of this and had limited enforcement capacity despite absolute authority), though the army was present near the borders, there were few police, and other public services were non-existent. Local tribal communities were left more-or-less untouched, so long as crime reports remained acceptable. At the same time, more troubled regions were brought under FCR—including large parts of the Balochistan and Sindh provinces.

Over the next half-century, FCR changed very little. Besides extending it to a few additional regions, the legal systems and lack of other institutions remained fixed. The British had found an acceptable solution in dealing with these areas.

3.2.2 The Frontier Crimes Regulation Since Independence (1947-2012)

Perhaps surprisingly, after independence FCR was not revoked from most of modern-day Pakistan; the language of the regulation was left intact for over half of a century. Political agents were still appointed, now by the head of the Punjab Province. Cases still went to Jirgas. In fact, several years after the country's independence, FCR was extended to including additional parts of Balochistan and, briefly, new areas in Punjab and Sindh. It was only over the course of several decades that it was slowly rolled back to the tribal areas which remain under FCR today. We detail these geographic changes in Section 3.3 below.

Throughout this time period, FCR stopped being about controlling criminal activity and became more a choice to not extend the new government to tribal areas. For example, the debate in recent decades has shifted much more towards representation, as it was not until 1997 that Pakistanis in FCR regions were even granted representation in the national legislature. Party-based elections were only introduced to areas under the FCR in 2013, decades after the rest of the country.

3.3 Data

3.3.1 FCR Application, 1901-2012

In order to understand both the British and later Pakistan's decisions to apply FCR to and continue to maintain FCR in large parts of Pakistan, we use primary legal documents to create a dataset of when and where FCR has applied between 1901 and 2012 for all 403 sub-districts (tehsils) in Pakistan. Basic summary stats are presented in Table 3.1 and in Figure 3.1. The years selected in the table and figure were intentional. They represent all of the years in which there have been changes in FCR status of at least one sub-district, in addition to 2012 (or present as it has not changed since then). The first two years demonstrate that there was very little change in FCR application between 1901 and Pakistan's independence from the British in 1947. The following six years follow the changes that occurred before and after the Green Revolution. In 1965, the biggest roll-back in FCR thus far occurred. This roll-back will provide the primary variation for our differences-in-differences analysis, which we will discuss below. The choice of 2012 demonstrates that FCR application has not changed since 1978.

3.3.2 Crop Suitability and the Green Revolution

For a time-invariant measure of potential crop yields, we utilize crop suitability data from the Food and Agriculture Organization of the United Nations (FAO, 2012). The FAO provides us with sub-district level indices of agro-climactical suitability for a variety of crops. We focus on wheat which was by far the most common crop in Pakistan around the time of the Green Revolution and the crop that would overwhelmingly benefit from the new technologies. The FAO indices are based on factors such as location-specific geography, rainfall, and temperature over the period 1961-1990. Our measure of crop suitability is the average of these FAO indices across different potential irrigation levels at low input.

Figure 3.2 shows the extent of geographic variation in crop suitability for wheat. While most of Pakistan falls in the medium to not suitable categories, there is a fair amount of geographic variation, especially in areas that at one point had or have FCR.[6]

Though the data used to create these FAO indices include more recent weather information than many of the years in our analysis, we believe that the cross-sectional variation applies across this time period given that the geographic are fixed and that rainfall and temperature are very slow to change.

Importantly, we have also documented the point at which the Green Revolution first began in Pakistan—1965. These changes were driven by the technological changes in wheat production, which was the most important Green Revolution Crop in Pakistan. And with wheat the key changes were not in terms of inputs. Rather, the key change was the introduction of new high-yielding varieties first introduced

[6]Note that the FAO crop suitability data is provided in raster images with various resolutions depending on the crop. Sub-district-level means for each input level are extracted from each raster images, and then these means are averaged to form a single index for each crop.

in Punjab in 1965.[7] In Western Pakistan, wheat production increased by 79 percent from 1966 to 1969, with a peak growth rate of agricultural output of 15 percent during fiscal 1967-68 (Child and Kaneda, 1975).[8]

The Green Revolution in South Asia was characterized by increased crop yields among the staple crops. With wheat there were few required changes in input technologies, labor to capital ratios, or irrigation. We will therefore consider the Green Revolution to mitigate the importance of crop suitability for wheat.[9] This is consistent with Foster and Rosenzweig (1996) and with Child and Kaneda (1975).[10]

3.4 Empirical Approach

We conduct two complementary analyses of the choice to apply, and then maintain, FCR provision in regions of Pakistan. First, we correlate fixed, sub-district-level characteristics, including crop suitability, with the initial decision that the British made to select roughly half of Pakistan for FCR in 1901. Second, we exploit the differential impact of the Green Revolution by crop suitability to understand Pakistan's decisions to roll FCR back across parts of the country throughout the 1960s and 1970s.

[7]See Dowswell (1989). Using similar data, the International Maize and What Improvement Center (CIMMYT) reports that the 118156 wheat variety, the basis for the most important Green Revolution varieties, was first released in 1966 (Lantican et al., 2012).

[8]We are currently collecting data on when Green Revolution varietals were introduced in each sub-district which will make our estimates of the onset of the Green Revolution more precise.

[9]This is different from Southeast Asia where the introduction of new rice varieties effected both input requirements (more fertilizer) as well as the modes of cultivation and distribution of returns (see e.g. Scott, 1977).

[10]Note that we are unable to provide district-specific trends in Green Revolution take-up for Pakistan as Foster and Rosenzweig (1996) do for India due to a lack of available data.

3.4.1 Initial FCR Application in 1901

For our first analysis, we will use a simple empirical specification:

$$\begin{aligned} \text{FCR\_applied\_1901}_d = \alpha + \beta_1\, \text{Sub-district\_area} + \beta_2\, \text{Ruggedness}_d \\ + \beta_3\, \text{Distance\_to\_capital}_d + \beta_4\, \text{Crop\_suitability}_d + \Gamma_d + \epsilon_d \end{aligned} \tag{3.1}$$

Where $\text{FCR\_applied\_1901}_d$ is a dummy for whether FCR was initially applied to sub-district d in the 1901 FCR legislation, $\text{Sub-district\_area}_d$ is is the area of sub-district d in thousands of square kilometers, Ruggedness_d is a sub-district measure of terrain roughness (SD of height above sea level), $\text{Distance\_to\_capital}_d$ is the distance, in 1000s of kilometers, from the centroid of each sub-district to the capital through the late 1960s (Karachi), $\text{Crop\_suitability}_d$ is a sub-district's crop suitability measure, and Γ_d are sub-district covariates. Note that FCR was originally applied at the district level, so we cluster the standard errors by district. We leave the specification at the sub-district level, however, to avoid having to aggregate up the geo-specific measure any more than has already been done.

This analysis will give us a correlation. What is informative is that while sub-district geographic characteristics that proxy for productivity at the time of territorial demarcation (which happened over the latter half of the 19th century), the challenge of exerting control, and transportation costs all correlate in the expected direction with FCR application, FCR application is conditionally independent of crop suitability for wheat. We proxy for initial productivity with sub-district area because administrative units were sized to capture similar populations during the initial demarcation in the mid-19th century and more productive places were more densely populated at the time. We therefore expect productive areas to have had physically smaller administrative units. We proxy for the challenge of exerting control with the standard deviation of elevation. It is well established that rougher terrain is harder to police given modern military technologies and this was certainly true in a time before mechanized transportation. We also proxy for transportation costs with the distance to the Karachi, the main port at the time for areas that would become Pakistan, because wheat was an important export crop. The value of controlling territory where it was produced was therefore likely related to the costs of moving it to market. We do not have pre-1901 data to control for potential omitted variables such as differential time trends in productivity, or for specific time-invariant covariates of a sub-district. As such, we will only consider results from this analysis as suggestive. Note we are in the process of coding up tax revenue data to control for one large potential omitted variable.

3.4.2 FCR Application and the Green Revolution

For our second analysis we exploit pre-existing cross-sectional variation in the marginal impact of Green Revolution wheat varieties on productivity with an exogenously timed technological change (the introduction of those varieties) to identify incentives for rolling back the FCR. Our primary specification will be as follows:

$$\begin{aligned}\text{FCR\_applied}_{dt} &= \alpha + \beta_1 \,\text{Crop\_suitability}_d + \beta_2 \,\text{Post\_GR}_t + \\ &\quad \text{Post\_GR\_Crop\_suitability}_{dt} + \delta_d + \delta_t + \epsilon_{dt} \\ &\quad \text{for sub-district } d \in \{\text{ever had fcr}\} \\ &\quad \text{for year } t \in \{1947, 1957, 1962, 1964, 1965, 1974, 1978\}\end{aligned} \tag{3.2}$$

Here FCR_applied_{dt} is a dummy for whether FCR continued to apply to sub-district d in year t, $\text{Crop\_suitability}_d$ is our crop suitability measure of sub-district d, and $\text{Post\_GR\_Crop\_suitability}_{dt}$ is the linear interaction of the the two terms. δ_d and δ_t are sub-district and year fixed effects. Note that we will not be able to separately identify β_1 from sub-district fixed effects.

Analysis for Equation 3.2 is limited to sub-districts in Pakistan that ever had FCR and to years $t \in \{1947, 1957, 1962, 1964, 1965, 1974, 1978\}$. The latter limitation is to all the years in which one or more sub-districts changed FCR application, within 20 years of the Green Revolution.[11] We limit to these years as an event study of sorts, assuming that there was enough of a political cost to changing the FCR legislation that it could not be done continuously. This approach matches the historical record in that decisions to remove sub-districts from the law happened episodically and in groups. There are two more extreme alternatives: (i) leave the data at the yearly level and run the same specification; or (ii) collapse the data down to two observations for each sub-district and run a simple difference of means between pre and post the Green Revolution. We see our specification as superior to (i) because it will not over-emphasize the many zeros that likely did not represent real decisions and to (ii) because it allows for a more accurate accounting for variation across time.[12]

With sub-district and year fixed effects, and with a differences-in-differences estimator, we will consider this analysis to capture the causal differential impact of the Green Revolution, or more generally of an exogenously timed change in a sub-district's agricultural land value, on the choice by the Pakistani government to maintain or remove FCR.[13] For our identification strategy to hold, we need that there were no time-varying omitted variables that differentially impacted sub-districts before and after 1965. In other words, we need that there were no other major changes other

[11]And more or less within a much larger window considering the little change in FCR between 1901 and 1947 and the no change in FCR after 1978.

[12]Note that if we take the conservative approach and run analysis on data for all 20 years before and after the Green Revolution as in (i), we obtain coefficients with 1/3 to 1/2 of the magnitude and the same level of significance. These are still very meaningful magnitudes. Results available upon request.

[13]In the context of our theory we have a one-time shock which impacts I but not k or k_T.

than the Green Revolution happening at or around 1965 that had differential impacts on FCR application by crop suitability. We will discuss other major potential changes in the context of our results below.

3.5 Results

This section presents results from two complementary analyses of the choice to apply, and then maintain, FCR provision in regions of Pakistan. First, we correlate fixed, sub-district-level crop suitability with the initial decision that the British made to select roughly half of Pakistan for FCR in 1901. Second, we exploit the differential impact of the Green Revolution by crop suitability to understand Pakistan's decisions to roll FCR back across parts of the country throughout the 1960s and 1970s.

3.5.1 Initial FCR Application in 1901

Table 3.2 presents results for this analysis. In-line with our model, column (1) shows that the British applied FCR in less productive places (under our assumption that sub-district area was negatively correlated with productivity), column (2) shows that that places which were more costly to tax due to rough terrain, as proxied by height above sea level, were more likely to be in the FCR, and column (3) shows that places with higher transportation costs (and thus lower revenue potential given productivity) were more likely to be included. Column (4) shoes that there was a positive correlation between wheat crop suitability and initial FCR application, but Column (5) then shows that once our other factors are accounted for the sub-district suitability for wheat is uncorrelated with initial FCR application.

Thus we argue that initial assignment to the FCR is plausibly exogenous to wheat crop suitability once we condition for proxies for local productivity and the cost of imposing state institutions.

3.5.2 FCR Application and the Green Revolution

Table 3.3 presents results for our second analysis—exploiting the differential impact of the Green Revolution by crop suitability to understand the integration of peripheral areas into the state throughout the 1960s and 1970s. We first present a simple correlation of sub-district crop suitability and FCR application across the years in this analysis with and without year fixed effects. In columns (1) and (2), we see that there is no clear correlation between sub-district wheat suitability and FCR status when we do not use our Green Revolution instrument. This is not surprising as our previous results from 1901 suggest that wheat suitability might not have been an important predictor of FCR application initially, and there were very few changes in FCR before the Green Revolution. It is also consistent with the fact that the Green Revolution cause wheat suitability to matter differentially at a discrete point in time.

Second, we present a differences-in-differences specification with year and sub-district fixed effects, our preferred specification. We can see that crop suitability differentially positively predicts FCR's continued application after the Green Revolution relative to before by about 10 percent in column (3). We then divide crop suitability at the median in column (4) to show that below median crop suitability sub-districts are *less* likely to have FCR retained after the Green Revolution. This point estimate is substantively large — 31.6 percentage points. These results are confirmed visually in Figure 3.3, where we group sub-districts into above and below median crop suitability and show mean FCR application levels over time for all those sub-districts that ever had FCR. We see that after the Green Revolution low suitability districts became much less likely to have FCR maintained.

This result is consistent with the fact that the Green Revolution mitigated the importance of crop suitability. As mentioned above, the Green Revolution is characterized by increased crop yields among the staple crops of South Asia with little to no required changes in input technologies, labor to capital ratios, or irrigation. Thus places that were once harder to farm became relatively easier, causing lower-suitability sub-districts to 'catch-up' to other districts in potential revenue extraction. Thus lower-suitability districts were more likely to switch from expected revenue negative to positive as a result of the Green Revolution, and these districts were relatively more likely to have their FCR application removed.[14]

Our results suggest a fairly large magnitude of an effect as well. Using our preferred specification in Table 3.3 column (3), we see that a one unit increase in crop suitability, from say 'medium' to 'good,' is associated with a 10 percentage points differential increase in the likelihood that FCR continues to apply to a sub-district following the Green Revolution.

3.5.3 Robustness Checks

We present three robustness checks to our Green Revolution result. First, we conduct a placebo check in which we add to our main specification the interaction of a dummy for post Green Revolution with crop suitability levels for other crops that were not impacted by the Green Revolution in Pakistan during this time period (rice became important in the 1980s), including gram (the second most important crop in Pakistan before the Green Revolution by cultivated area), and an average over gram, cotton, and rice. Results are presented in Appendix Table 3.6. Note we do not show cotton or rice separately. For cotton, this is because it is 97 percent correlated with wheat suitability in our sample. For rice, it is because there is no variation in rice suitability in all of Balochistan (it is zero throughout). In columns (2) and (3),

[14] This is consistent with Foster and Rosenzweig (1996) and with Child and Kaneda (1975).

our results remain completely driven by wheat crop suitability. This assures us that our results are not driven by some correlate of wheat suitable sub-districts that has nothing to do with potential land revenue—if such was the case it seems unlikely such a spurious result would exist with crop suitability for wheat but not the second most important crop or an average across three important crops.

Second, we conduct a placebo check in which we vary the year in which the Green Revolution supposedly took place. If there were pre-existing trends in low- relative to high-suitability sub-districts, such a placebo check should pick them up. Results are presented in Appendix Table 3.7. As you can see, we only get results when we use the true year of the Green Revolution in Pakistan. This is encouraging, and consistent with the pre-trends visible in Figure 3.3.

Third, most importantly, we turn to alternative changes that could have occurred in exactly 1965 in Pakistan that differentially affected some sub-districts over others in a way that is correlated with both wheat suitability and FCR application.[15] From the history of Pakistan around 1965, two plausible alternative stories arise. The first is Pakistan's on-going dispute with India over areas of northern Pakistan, including Kashmir and Gilgit-Baltistan. In 1965, there was a war between the two countries in this region which involved skirmishes in Kashmir (the Indo-Pakistani War of 1965). It is possible that this war constrained Pakistan's ability to roll back FCR from the northern region of the country in 1965. If this is the case, and the northern area is on average more suitable for wheat than Balochistan, where FCR was mainly rolled back in 1965, we could be obtaining spurious results. To ensure this is not the case, in Table 3.8, column (2), we limit our analysis to only Balochistan, the province for which there are major changes in FCR status in 1965. We see that, if anything, our result becomes stronger when limited to only Balochistan. This ensures that the

[15] Note we consider the Green Revolution to encapsulate all changes in crop technology at the time, so we aren't concerned about other simultaneous agricultural advances.

northern areas of Pakistan are not driving our results.

The second alternative story that we consider takes place within Balochistan. Balochistan has historically been dominated by two ethnic groups—the Balochis and the Pashtuns. The Pashtuns have lived in the northern part of the district and the Balochis in the Southern. It is possible that the President of Pakistan in 1965, General Ayub Khan, had different relationships with the two groups that might have affected FCR decisions, and in this time period the president was very active domestically. If the Pashtun or Balochi groups were systematically on more or less suitable land, we could be obtaining spurious results. To ensure this is not the case, in Table 3.8, column (3), we limit our analysis to only the historically Balochi sub-districts in Balochistan.[16] We find that our results generally hold (our standard errors get larger with the smaller sample but the main result remains significant at 10%). Thus, putting together our results from this table, even in the Balochi-dominated districts in Balochistan, we see the same relationship holding in which crop suitability differentially positively predicts FCR's continued application after the Green Revolution relative to before.

3.6 Conclusion

In this paper, we showed that the trajectory of state presence within the borders of modern day Pakistan is consistent with a model in which states extend governance to areas where the economic benefits of developing full institutions through taxation and resource extraction outweigh the costs of doing so. Using crop suitability data from the Food and Agriculture Organization of the United Nations, we show first that the choice by the British to apply FCR to over half of Pakistan in 1901 was conditionally uncorrelated with crop suitability. We then exploit the fact that

[16]Historically Balochi and Pashtun sub-districts were hand-coded using a map created by Dr. Michael Izady at www.Gulf2000.Columbia.edu/maps.shtml

Green Revolution had a greater marginal effect in areas of low crop suitability to understand Pakistan's selective roll-back of FCR throughout the 1960s and 1970s. We find that sub-districts more suitable to agriculture were more likely to see continued FCR application after the Green Revolution raised the relative value of less-suitable sub-districts.

Because the timing of the Green Revolution in Pakistan was exogenous to local politics we are able to isolate a plausibly causal effect of agricultural land value on FCR application. Our results suggest a large effect of land value on FCR application. Specifically, a one unit increase in crop suitability from 'medium' to 'good' increased a sub-district's probability of being left ungoverned by over twenty percent following the Green Revolution, relative to before. Though counter intuitive at first glance, this is actually consistent our hypothesis that the Green Revolution mitigated the importance of crop suitability and thus caused lower-suitability sub-districts to 'catch-up' to other districts in potential revenue extraction. Thus lower-suitability districts were more likely to switch from expected revenue negative to positive as a result of the Green Revolution, and these districts were relatively more likely to have FCR removed.

These results are important for at least four reasons. First, we provide microeconomic evidence on the importance of extractable land value, and technology-driven changes in land value, to the choice to govern land, supporting the hypothesis of a rich macroeconomic development literature when applied at the sub-national level. Second, we provide additional evidence on the importance of the Green Revolution in South Asia, not only in increasing land values and growth but in influencing the choice of the Pakistani government to govern (and Pakistani citizens to accept government in) large parts of the country that had thus far remained ungoverned. Third, we present microeconomic evidence in support of the idea of Fearon (2008) and Besley and Persson (2011) of *ungoverned-by-choice* space. Lastly, we provide heartening evidence that technological chance can lead to ungoverned spaces being folded into country's cores without civil war or serious violence. The parts of Pakistan that still have FCR today are, of course, the most resistant to government control, but so were many parts of the sub-districts that were brought into the government in the 1970s. Yet what was stopping the government from integrating them, and those living in these peripheral areas from consenting to integration without violence, at least in part, a simple cost-benefit calculation.

3.7 Acknowledgements

Chapter 3, in part, is currently being prepared for submission for publication of the material. Callen, Michael; Gulzar, Saad; Rezaee, Arman; Shapiro, Jacob.Shapiro acknowledges support from AFOSR, grant #FA9550-09-1-0314.

3.8 Chapter 3 Appendix

3.8.1 Tables and Figures

Table 3.1: FCR Application Summary Statistics

	% of Sub-districts under FCR	% area under FCR (km^2)
Year:		
1901	42.93	52.08
1947	42.43	50.07
1957	43.42	58.15
1962	46.65	59.66
1964	34.00	52.12
1965	23.57	21.77
1974	15.63	10.21
1978	11.91	02.97
2012	11.91	02.97

Notes: Percentage sub-districts (tehsils) under FCR based on a total of 403 sub-districts. Area under FCR based on a total area of 872,027 square kilometers.

Table 3.2: Crop Suitability and Initial FCR Application

	FCR applied in 1901 (=1)				
	(1)	(2)	(3)	(4)	(5)
Sub-district Area (Square KM / 1000)	0.019**				0.016*
	(0.009)				(0.009)
Sub-district SD of height above sea level (FT / 100)		0.023***			0.024***
		(0.006)			(0.006)
Distance to Capital (KM / 1000)			0.351***		-0.088
			(0.107)		(0.165)
Sub-district Wheat Crop Suitability				0.061*	0.082**
				(0.031)	(0.038)
Mean of dependent variable	0.429	0.429	0.429	0.429	0.429
# Observations	403	403	403	403	403
# Clusters	129	129	129	129	129
R-Squared	0.016	0.145	0.074	0.034	0.192

Notes: Unit of observation is the sub-district (tehsil). $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard errors clustered at the district level reported in parentheses. Crop suitability scores are as follows: 0 is not suitable, 1 is very marginal, 2 is marginal, 3 is moderate, 4 is medium, 5 is good, 6 is high, and 7 is very high. Karachi was the capital at independence and the main export port in the late-1900s.

Table 3.3: Crop Suitability and FCR Application Before and After the Green Revolution

	FCR applied (=1)			
	(1)	(2)	(3)	(4)
Sub-district Wheat Suitability	0.029	0.029		
	(0.037)	(0.037)		
Post Green Revolution (=1)			-0.741***	-0.510***
			(0.076)	(0.095)
Wheat Suitability * Post Green Revolution			0.096***	
			(0.033)	
Below Median Wheat Suitability (=1) * Post				-0.270***
				(0.099)
Mean of dependent variable	0.952	0.952	0.952	0.952
# Observations	1421	1421	1421	1421
# Clusters	74	74	74	74
R-Squared	0.006	0.291	0.454	0.454
Year FEs?	NO	YES	YES	YES
Sub-district FEs?	NO	NO	YES	YES

Notes :$*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard errors clustered at the district level reported in parentheses. Crop suitability scores are as follows: 0 is not suitable, 1 is very marginal, 2 is marginal, 3 is moderate, 4 is medium, 5 is good, 6 is high, and 7 is very high. Post Green Revolution is a dummy for years after 1963. Years in analysis limited to those years where any sub-district had FCR removed—1922,1937,1947,1956,1963,1964,1971,1973,1977.

Table 3.4: Take-up of Green Revolution Crop Complements

	Difference between 1972 and 1960 values					Farms with Tractors '72 – Ploughs '60
	Cropped Area	Wheat Area	Irrigated Area	Cropped Area with Fertilizer	Farms using Fertilizer	
	(1)	(2)	(3)	(4)	(5)	(6)
Below Median Wheat Suitability (=1)	127238	44502	63344	79682**	11070***	6582**
	(135862)	(47821)	(80983)	(35189)	(3857)	(2791)
Mean of dependent variable	67157	13299	28653	52014	-1116	2653
# Observations	38	38	35	34	34	25
# Clusters	27	27	25	25	25	17
R-Squared	0.066	0.078	0.065	0.275	0.315	0.129

Notes :*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors clustered at the district level reported in parentheses. Crop suitability scores are as follows: 0 is not suitable, 1 is very marginal, 2 is marginal, 3 is moderate, 4 is medium, 5 is good, 6 is high, and 7 is very high. All areas are in Acres. Years in analysis limited to those years where any sub-district had FCR removed—1922,1937,1947,1956,1963,1964,1971,1973,1977.

Table 3.5: IV Results on Literacy from the PKDHS

	Literate (=1)		
	(1)	(2)	(3)
Sub-district under FCR (=1)	-0.190	-0.110	-0.244*
	(0.116)	(0.090)	(0.142)
Post Green Revolution (=1)	-0.150	-0.057	-0.169
	(0.121)	(0.094)	(0.144)
Sub-district Wheat Suitability	0.029	-0.012	-0.001
	(0.018)	(0.011)	(0.011)
HH Wealth Index		0.226***	0.141***
		(0.014)	(0.046)
Mean of dependent variable	0.181	0.181	0.048
# Observations	1259	1259	627
# Clusters	42	42	39
R-Squared	0.030	0.302	.
First-stage F	24.200	23.755	13.652
Sample	All	All	<=p50 WI

Notes :*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors clustered at the district level reported in parentheses. All regressions include year and province fixed effects, and are limited to households with heads born within ten years of 1965 that have not migrated. District under FCR is instrumented with the interaction of Post Green Revolution (=1) and District Wheat Suitability.

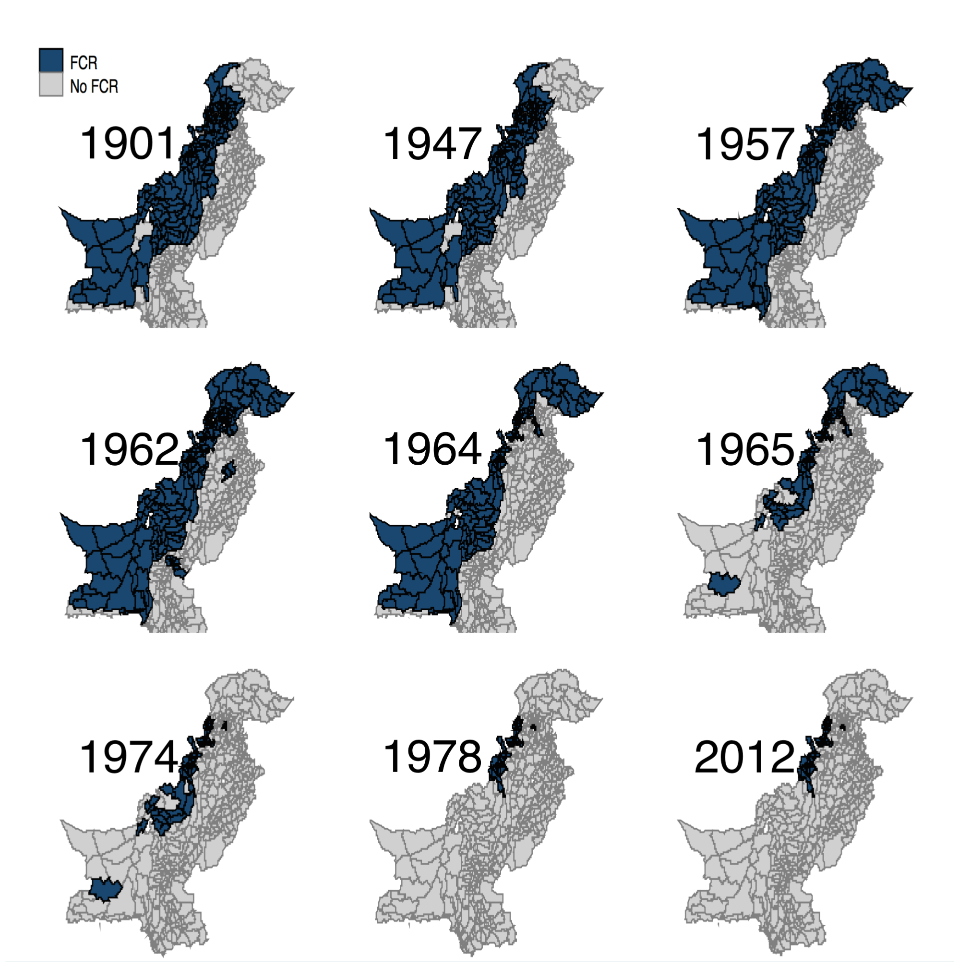

Figure 3.1: FCR Application Over Time

Sub-district (tehsil) boundaries marked. White sub-districts are those for which we do not have data, due to changes in sub-district boundaries between 1901 and 2012.

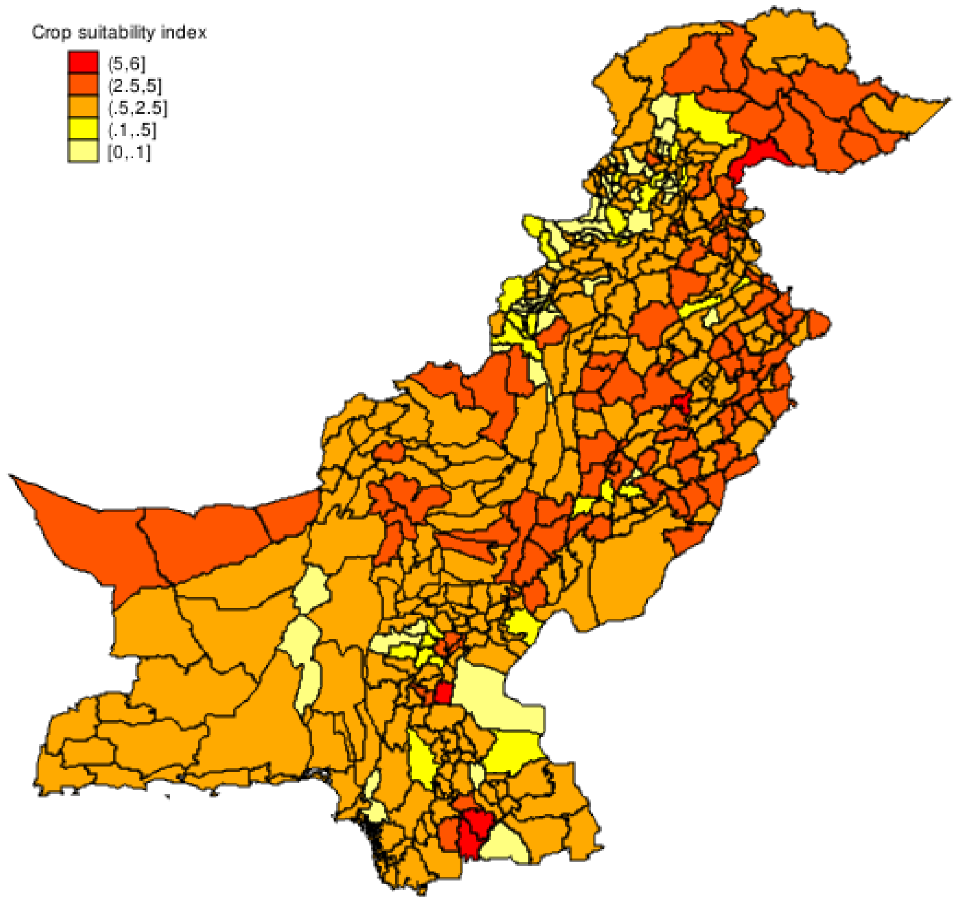

Figure 3.2: Crop Suitability

Sub-district (tehsil) boundaries marked. Crop suitability scores are as follows: 0 is not suitable, 1 is very marginal, 2 is marginal, 3 is moderate, 4 is medium, 5 is good, 6 is high, and 7 is very high. Data from FAO, 2012.

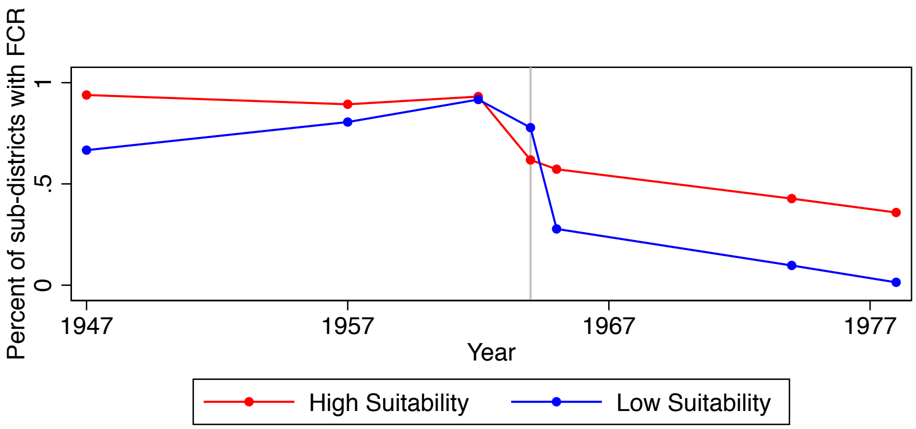

Figure 3.3: FCR Application Over Time by Crop Suitability

Points show the mean sub-district FCR application dummy values in years used in analaysis within above (high) and below (low) medium wheat crop suitability bins. Lines are fitted using locally weighted scatterplot smoothing.

3.8.2 Appendix Tables and Figures

Table 3.6: Robustness Check 1—Other Crop Suitability Placebos

	FCR maintained (=1)		
	(1)	(2)	(3)
Wheat Crop Suitability * Post Green Revolution	0.096***	0.087	0.147*
	(0.033)	(0.074)	(0.082)
Gram Crop Suitability * Post Green Revolution		0.011	
		(0.083)	
Other Crop Average Crop Suitability * Post Green Revolution			-0.090
			(0.125)
Mean of dependent variable	0.952	0.952	0.952
# Observations	1421	1421	1421
# Clusters	74	74	74
R-Squared	0.454	0.454	0.455
Year FEs?	YES	YES	YES
Sub-district FEs?	YES	YES	YES

Notes :$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the district level reported in parentheses. Other crop suitability is the average of gram, cotton, and rice crop suitability.

Table 3.7: Robustness Check 2—Year Placebos

	FCR maintained (=1)			
	(1)	(2)	(3)	(4)
Wheat Crop Suitability * Post Green Revolution	0.096***			
	(0.033)			
Wheat Crop Suitability * Post 1957		-0.063		
		(0.046)		
Wheat Crop Suitability * Post 1962			-0.008	
			(0.033)	
Wheat Crop Suitability * Post 1964				0.031
				(0.035)
Mean of dependent variable	0.952	0.952	0.952	0.952
# Observations	1421	1421	1421	1421
# Clusters	74	74	74	74
R-Squared	0.454	0.433	0.428	0.430
Year FEs?	YES	YES	YES	YES
Sub-district FEs?	YES	YES	YES	YES

Notes :$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the district level reported in parentheses.

Table 3.8: Robustness Check 3—Focus on Balochistan

	FCR maintained (=1)		
	(1)	(2)	(3)
Wheat Suitability * Post Green Revolution	0.096***	0.198**	0.190*
	(0.033)	(0.083)	(0.100)
Mean of dependent variable	0.952	0.842	0.799
# Observations	1421	420	315
# Clusters	74	25	18
R-Squared	0.454	0.776	0.845
Year FEs?	YES	YES	YES
Sub-district FEs?	YES	YES	YES
Sample	All	Balochistan Only	Non-Pashtun Balochistan

Notes :$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the district level reported in parentheses.

Bibliography

Acemoglu, Daron, James A Robinson, and Rafael J Santos, "The monopoly of violence: Evidence from colombia," *Journal of the European Economic Association*, 2013, *11* (s1), 5–44.

_ , Simon Johnson, and James A Robinson, "The colonial origins of comparative development: an empirical investigation," *American Economic Review*, 2001, *91* (5), 1369–1401.

Aker, Jenny C, "Information from markets near and far: Mobile phones and agricultural markets in Niger," *American Economic Journal: Applied Economics*, 2010, *2* (3), 46–59.

Alesina, Alberto and Enrico Spolaore, "On the Number and Size of Nations," *Quarterly Journal of Economics*, 1997, *112* (4), 1027–1056.

_ and _ , *The Size of Nations*, MIT Press, 2003.

_ and Romain Wacziarg, "Openness, Country Size and Government," *Journal of Public Economics*, 1998, *69*, 305–321.

Almlund, Mathilde, Angela Lee Duckworth, James J Heckman, and Tim D Kautz, "Personality psychology and economics," Technical Report, National Bureau of Economic Research 2011.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja, "Report cards: The impact of providing school and child test scores on educational markets," 2014.

Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack, "No margin, no mission? A Field Experiment on Incentives for Public Service Delivery," *Journal of Public Economics*, Forthcoming, *8834*.

_ , _ , and Scott S Lee, "Do-gooders and go-getters: career incentives, selection, and performance in public service delivery," 2014.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger, "Working Over Time: Dynamic Inconsistency in Real Effort Tasks," *The Quarterly Journal of Economics*, Forthcoming.

Bandiera, Oriana, Andrea Prat, and Tommaso Valletti, "Active and Passive Waste in Government Spending: Evidence from a Policy Experiment," *American Economic Review*, 2009, *99* (4), 1278–1308.

Bapat, Navin, "Transnational Terrorism, U.S. Military Aid, and the Incentive to Misrepresent," *Journal of Peace Research*, 2011, *48* (3), 315–335.

Barrick, Murray R. and Michael K. Mount, "The Big Five Personality Dimensions and Job Performance: A Meta-Analysis," *Personnel Psychology*, 1991, *44* (1), 1–26.

Bazerman, Max and Don A Moore, *Judgment in Managerial Decision Making, 8th Edition*, Wiley & Sons, 2012.

Benabou, Roland and Jean Tirole, "Intrinsic and extrinsic motivation," *The Review of Economic Studies*, 2003, *70* (3), 489–520.

Benjamini, Yoav and Yosef Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, pp. 289–300.

Berman, Eli, Jacob N Shapiro, and Joseph H Felter, "Can hearts and minds be bought? The economics of counterinsurgency in Iraq," *Journal of Political Economy*, 2011, *119* (4), 766–819.

Bertrand, Marianne and Antoinette Schoar, "Managing with Style: The Effect of Managers on Firm Policies," *Quarterly Journal of Economics*, 2003, *CXVIII*, 1169–1208.

Besley, Timothy and Maitreesh Ghatak, "Competition and incentives with motivated agents," *American Economic Review*, 2005, *95* (3), 616–636.

Blair, Graeme, C Christine Fair, Neil Malhotra, and Jacob N Shapiro, "Poverty and support for militant politics: Evidence from Pakistan," *American Journal of Political Science*, 2013, *57* (1), 30–48.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan, "Reducing crime and violence: Experimental evidence on adult noncognitive investments in Liberia," 2015. Working paper.

Borghans, Lex, Angela Lee Duckworth, James J. Heckman, and Bas ter Weel, "The Economics and Psychology of Personality Traits," *The Journal of Human Resources*, 2008, *XLIII* (4), 973–1059.

Bowles, Samuel, Herbert Gintis, and Melissa Osborne, "The determinants of earnings: A behavioral approach," *Journal of Economic Literature*, 2001, pp. 1137–1176.

Callen, Michael, Saad Gulzar, Ali Hasanain, and Yasir Khan, "The Political Economy of Public Employee Absence: Experimental Evidence from Pakistan," 2013.

_ , _ , _ , _ , and Arman Rezaee, "Personalities and Public Sector Performance: Evidence from a Health Experiment in Pakistan," Technical Report, National Bureau of Economic Research 2015.

Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F Halsey Rogers, "Missing in action: teacher and health worker absence in developing countries," *The Journal of Economic Perspectives*, 2006, *20* (1), 91–116.

Chetty, Raj, John N Friedman, and Jonah E Rockoff, "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates," *American Economic Review*, 2014.

_ , _ , and _ , "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood," *American Economic Review*, 2014.

Child, Frank C and Hiromitsu Kaneda, "Links to the green revolution: a study of small-scale, agriculturally related industry in the Pakistan Punjab," *Economic Development and Cultural Change*, 1975, pp. 249–275.

Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi, "Strengthening State Capabilities: The Role of Financial Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *Quarterly Journal of Economics*, 2013, *forthcoming.*

Deaton, Angus, "Instruments, Randomization, and Learning about Development," *Journal of Economic Literature*, 2010, *48* (2), 424–55.

Diamond, Jared M, *Guns, germs and steel: a short history of everybody for the last 13,000 years*, Random House, 1998.

Dowswell, Christopher R, *Wheat research and development in Pakistan*, CIMMYT, 1989.

Duflo, Esther, Rema Hanna, and Stephen P Ryan, "Incentives work: Getting teachers to come to school," *The American Economic Review*, 2012, pp. 1241–1278.

Easterly, William and Sergio Rebelo, "Fiscal Policy and Economic Growth: An Empirical Investigation," *Journal of Monetary Economics*, 1993, *32* (417-458).

Fafchamps, Marcel and Bart Minten, "Impact of sms-based agricultural information on indian farmers," *The World Bank Economic Review*, 2012, *26* (3), 383–414.

Fearon, James D and David D Laitin, "Ethnicity, insurgency, and civil war," *American political science review*, 2003, *97* (01), 75–90.

Felter, Joseph H., "Aligning Incentives to Combat Terror," in Rohan Gunaratna, ed., *Combating Terrorism (Regionalism and National Security*, Marshall Cavendish Academic, 2006.

Ferraz, Claudio and Frederico Finan, "Electoral accountability and corruption: Evidence from the audits of local governments," *American Economic Review*, June 2011, *101* (4), 1274–1311.

Food and Agricultural Organization of the United Nations, "Global Agro-Ecological Zones (GAEZ v3.0)," 2012. Available at http://gaez.fao.org.

Foster, Andrew D and Mark R Rosenzweig, "Technical change and human-capital returns and investments: evidence from the green revolution," *The American economic review*, 1996, pp. 931–953.

Friedman, David, "A Theory of the Size and Shape of Nations," *Journal of Political Economy*, 1977, *85* (1), 59–77.

Gallup, John Luke, Jeffrey D Sachs, and Andrew D Mellinger, "Geography and economic development," *International regional science review*, 1999, *22* (2), 179–232.

Gatewood, Robert, Hubert Feild, and Murray Barrick, *Human resource selection*, Cengage Learning, 2010.

Grossman, Herschel I and Suk Jae Noh, "A Theory of Kleptocracy with Probabilistic Survival and Reputation," *Economics & Politics*, 1990, *2* (2), 157–171.

_ and _ , "Proprietary public finance and economic welfare," *Journal of Public Economics*, 1994, *53* (2), 187–204.

Groth-Marnat, Gary, *Handbook of psychological assessment*, John Wiley & Sons, 2009.

Hanushek, Eric A, "The economic value of higher teacher quality," *Economics of Education Review*, 2011, *30* (3), 466–479.

_ and Steven G Rivkin, "The distribution of teacher quality and implications for policy," *Annu. Rev. Econ.*, 2012, *4* (1), 131–157.

Heckman, James J., "Integrating Personality Psychology into Economics," 2011, (NBER WP #17378).

_ , Jora Stixrud, and Sergio Urzua, "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 2006, *24* (3), 411–482.

Henderson, Ralph H and T Sundaresan, "Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method.," *Bulletin of the World Health Organization*, 1982, *60* (2), 253.

Hogan, Joyce and Brent Holland, "Using Theory to Evaluate Personality and Job-Performance Relations: A Socioanalytic Perspective," *Journal of Applied Psychology*, 2003, *88* (1), 100–112.

i Miquel, Gerard Padro and Pierre Yared, "The political economy of indirect control," *Quarterly Journal of Economics*, May 2012, *127* (2), 947–1015.

Jensen, Robert, "The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector," *The Quarterly Journal of Economics*, 2007, pp. 879–924.

Jin, Ginger Zhe and Phillip Leslie, "The effect of information on product quality: Evidence from restaurant hygiene grade cards," *The Quarterly Journal of Economics*, May 2003.

John, Oliver P., Laura P. Naumann, and Christopher J. Soto, "Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues," in "Handbook of personality: Theory and research," The Guilford Press, 2008, chapter 4.

Johnson, W. Bruce, Robert Magee, Nandu Nagarajan, and Harry Newman, "An Analysis of the Stock Price Reaction to Sudden Executive Deaths," *Journal of Accounting and Economics*, 1985, *7*, 151–174.

Jones, Benjamin F. and Benjamin A. Olken, "Do Leaders Matter? National Leadership and Growth Since World War II," *Quarterly Journal of Economics*, 2005, *120* (3), 835–864.

Kaplan, Robert M. and Dennis P. Saccuzzo, *Psychological Testing: Principles, Applications, and Issues*, Pacific Grove, Calif.: Brooks/Cole Pub. Co., 1997.

Kautz, Tim D, James J. Heckman, Ron Diris, Bas ter Weel, and Lex Borghans, "Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success," 2014, (NBER WP #20749).

Klinger, Bailey, Asim Ijaz Khwaja, and Carlos del Carpio, *Enterprising Psychometrics and Poverty Reduction*, Springer, 2013.

Kuhn, Michael A, "Curing the Calorie Crunch: The Effect of EBT on Household Present-Bias," 2013.

Lantican, M.A., H.J. Dubin, and M.L. Morris, International Maize and Wheat Improvement Center (CIMMYT), 2012.

Malmendier, Ulrike, Geoffrey Tate, and Jon Yan, "Overconfidence and Early-Life Experiences: The Effect of Managerial Traits on Corporate Financial Policies," *The Journal of Finance*, 2011, *66* (5), 1687–1733.

McArthur, John W and Jeffrey D Sachs, "Institutions and geography: comment on Acemoglu, Johnson and Robinson (2000)," Technical Report, National bureau of economic research 2001.

Miguel, Edward, C Camerer, K Casey, J Cohen, KM Esterling, A Gerber, R Glennerster, DP Green, M Humphreys, G Imbens et al., "Promoting transparency in social science research," *Science*, 2014, *343* (6166), 30–31.

Mitra, Sandip, Dilip Mookherjee, Maximo Torero, and Sujata Visaria, "Asymmetric information and middleman margins: An experiment with west bengal potato farmers," 2014. Working paper.

National Institute of Population Studies, *Pakistan Demographic and Health Survey 2012-13*, National Institute of Population Studies, 2013.

Nichols, Robert, *The Frontier Crimes Regulation: A History in Documents*, Oxford University Press, 2013.

Nunn, Nathan and Diego Puga, "Ruggedness: The blessing of bad geography in Africa," *Review of Economics and Statistics*, 2012, *94* (1), 20–36.

Nyhus, Ellen K. and Empar Pons, "The Effects of Personality on Earnings," *Journal of Economic Psychology*, 2005, *26* (3), 363–384.

Olken, Benjamin A. and Rohini Pande, "Corruption in Developing Countries," *Annual Review of Economics*, 2012, *4*, 479–509.

Perry, James L., "Measuring Public Service Motivation: An Assessment of Construct Reliability and Validity," *Journal of Public Administration Research and Theory*, 1996, *6* (1), 5–22.

_ **and Lois Recascino Wise**, "The Motivational Bases of Public Service," *Public Administration Review*, 1990, *50*, 367–73.

Petrovsky, Nicolai, "Does Public Service Motivation Predict Higher Public Service Performance? A Research Synthesis," 2009.

Rasul, Imran and Daniel Rogger, "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service," 2014.

Reinikka, Ritva and Jakob Svensson, "Local Capture: Evidence from a Central Government Transfer Program in Uganda," *The Quarterly Journal of Economics*, 2004, *119* (2), 679–705.

Roberts, Brent W., "Back to the Future: Personality and Assessment and Personality Development," *Journal of Research in Personality*, 2009, *43* (2), 137–145.

_ **, Kate E. Walton, and Wolfgang Viechtbauer**, "Patterns of Mean-Level Change in Personality Traits across the Life Course: A Meta-Analysis of Longitudinal Studies," *Psychological Bulletin*, 2006, *132* (1), 1–25.

Salgado, Jesus F., "The Five Factor Model of Personality and Job Performance in the The Five Factor Model of Personality and Job Performance in the European Community," *Journal of Applied Psychology*, 1997, *82* (1), 30–43.

Schmidt, Frank L and John E Hunter, "The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings.," *Psychological bulletin*, 1998, *124* (2), 262.

Scott, James C., *The Moral Economy of the Peasant: Rebellion and Subsistence in Southeast Asia*, Yale University Press, 1977.

Scott, James C, *The art of not being governed: An anarchist history of upland Southeast Asia*, Yale University Press, 2009.

Shapiro, Jesse M, "Is there a daily discount rate? Evidence from the food stamp nutrition cycle," *Journal of Public Economics*, 2005, *89* (2), 303–325.

Staiger, Douglas O and Jonah E Rockoff, "Searching for effective teachers with imperfect information," *The Journal of Economic Perspectives*, 2010, *24* (3), 97–117.

Svensson, Jakob and David Yanagizawa, "Getting prices right: the impact of the market information service in Uganda," *Journal of the European Economic Association*, 2009, *7* (2-3), 435–445.

Wild, Lena, Vikki Chambers, Maia King, and Daniel Harris, "Common Constraints and Incentive Problems in Service Delivery," Technical Report, Overseas Development Institute 2012.

World Bank, *World Development Report 2004: Making Services Work for the Poor*, World Bank, 2004.