Personalities and Public Sector Performance: Experimental Evidence from Pakistan*

Michael Callen† Saad Gulzar‡
Ali Hasanain§ Yasir Khan¶
Arman Rezaee‖

First Version: June 5, 2013
This Version: December 9, 2013

Abstract

We examine the relationship between policymaker personalities, job performance, and response to reforms in Punjab combining: (i) Big 5 personality and Perry Public Sector Motivation tests of the universe of health inspectors and senior health officials and a large and representative sample of doctors; (ii) measures of job performance from unannounced visits to health facilities; (iii) a randomized controlled evaluation of a novel smart phone monitoring technology; (iv) experimental manipulations of the presentation of data on doctor absence to senior health officials. Three results support the relevance of personalities for performance. First, Big 5 characteristics and Public Sector Motivation positively predict doctor attendance and negatively predict whether doctors collude with inspectors to falsify reports. Second, smart phone monitoring has the largest impact on health inspectors with high Big 5 characteristics—one SD higher health inspector Big 5 index is associated with a 27 percentage point differential increase in inspections due to increased monitoring. Last, senior health officials with high Big 5 characteristics are most likely to respond to a report of underperforming clinic as measured by improved subsequent performance at the facility—one SD higher senior health official Big 5 index is associated with an additional 40 percentage point reduction in doctor absence following underperforming facility flag in treatment districts.

Authors' Note: We thank Farasat Iqbal (Punjab Health Sector Reforms Project) for championing and implementing the project and, prinAsim Fayaz, and Zubair Bhatti (World Bank) for designing the program. Support is provided by the International Growth Centre (IGC) political economy program and the IGC Pakistan Country Office. We thank Eli Berman, Ali Cheema, Naved Hamid, Gordon Hanson, Asim Khwaja, Craig McIntosh, Ijaz Nabi, Christopher Woodruff and seminar participants at DFID, UC San Diego, and Growth Week at LSE for insightful comments. Excellent research assistance was provided by Muhammad Zia Mehmood and Haseeb Ali.

†University of California, Los Angeles. email: mjcallen@ucsd.edu
‡New York University. email: saad.gulzar@nyu.edu
§Lahore University of Management Sciences. email: hasanain@lums.edu.pk
¶International Growth Centre - Pakistan. email: yasir.khan@theigc.org
‖University of California, San Diego. email: arezaee@ucsd.edu

1 Introduction

Personality traits predict performance in many domains (Almlund et al., 2011; Borghans et al., 2008; Heckman, 2011). Reflecting this, firms, militaries, and governments in developed countries have long used psychometric measures to inform hiring, training, and promotion decisions (Kaplan and Saccuzzo, 1997). Because the poor rely primarily on governments for essential services, there is reason to study the role of personalities in public employee performance. Psychometrics may provide useful diagnostics and deeper insights into bureaucratic decision-making. In addition, recent research shows that the psychological profile of applicants to public jobs is largely determined by adjustable features of the position, most importantly the wage (Dal Bo et al., 2013). Personality traits are also malleable, providing a potential avenue for policy.[1] We study the role of personalities based on a comprehensive assessment of health officials in Punjab, Pakistan.[2]

Research in psychology and in economics points to a potential relationship between measures of non-cognitive traits and public employee performance. For example, Heckman et al. (2006) find that standardized adolescent measures of locus control and self-esteem (traits related to neuroticism, one of the Big 5 personality traits) predict adult earnings to a similar degree as cognitive ability. Specific to the Big 5 personality index, which we will use in this paper, Nyhus and Pons (2005) find using Dutch household data that wages are correlated with two of the Big 5 personality traits, emotional stability and conscientiousness.[3] Other

[1] In a meta-analysis, Roberts et al. (2006) examine 92 studies for patterns in the mean-level of Big 5 personality traits. The authors find that people increase in measures of social dominance (a facet of extraversion), conscientiousness, and emotional stability as they age, especially age 20 to 40. 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).

[2] According to 2008 population estimates, Punjab is the 9th largest subnational 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 in major health indicators, with a infant mortality rate of 77 per 1000 live births, for example. (See Provincial Annual Report, 2011, at health.punjab.gov.pk)

[3] Note that the authors also find strong heterogeneity in the returns to personality based on education group and gender. As our sample, explained in detail in Sections 2 and 3, is almost entirely male and of the same education level, we will not be able to make similar predictions.

meta-analyses find conscientiousness to be consistently predictive of earnings (Barrick and Mount, 1991; Salgado, 1997). Focused on job task performance rather than earnings, Hogan and Holland (2003) find in a meta-analysis that all five Big 5 measures positively predict performance on specific job criteria, and that the predictions become stronger as the job criteria become more specifics.

Economic studies find that leadership strongly impacts economic performance at both the firm level and the national level. 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 national level, Jones and Olken (2005) find, using deaths of leaders as exogenous variation, that leaders matter for a country's growth.

In this paper, we combine measures from personality psychology and two economic experiments to examine the relationship between personality measures and performance in health service provision. First, we correlate measures of Big 5 personality and Perry Public Sector Motivation with doctor attendance recorded through unannounced visits. Second, we examine whether these same measures predict systematic disagreement between official attendance records and our independently collected data, which provides evidence of data falsification. Third, using a randomized control trial, we examine whether the effect of a novel smartphone technology varies according to these measures. Last, we experimentally manipulate the presentation of actual absence data to senior health officials and investigate whether responses vary according to personality type.

We point to two central implications of these findings. First, absence among health providers is a serious policy issues in a large number of developing countries. The degree of correlation between personality measures, doctor attendance, and the responsiveness of senior officials to actionable data on absence suggest that substantial improvements can be

made by changing the profile of hired staff. Moreover, this may be achievable even in a system where incentives to attend work are weak. We view these results as complementary to the findings in Dal Bo et al. (2013). They show that increasing wages substantially improves the pool of applicants to public jobs, as measured by Big 5 and Perry Public Sector motivation measures. Our results indicate that workers with higher scores on these measures work more often and more effectively achieve their goals when given the same data. This is not always the case, however. In Section 6, we discuss what mechanisms might lead to personality being a strong predictor in certain settings and not others. Second, these results suggest that public worker heterogeneity has material implications for service delivery, even in settings where extrinsic incentives for performance are weak.

We document a strong relationship between personality measures and performance in health service provision. A one standard deviation increase in conscientiousness is associated with a 6 percentage point increase in attendance for doctors. This is a sizable impact; in our sample doctors are present at only 49 percent of visits. Measures of public sector motivation are also robustly associated with similar increases in predicted attendance. Similarly, a one standard deviation increase in conscientiousness of a doctor is associated with a 11 percentage point reduction in the rate of falsified official reports, more than 100% of the unconditional mean of falsified reports of 9 percent. We find no relationship between health inspector personality and performance, as measured by the rate of inspection of health facilities. We do, however, find that a one standard increase in the public service motivation aggregate z-score of health inspectors is associated with a 79 percent reduction in the rate of falsified reports, from 9.4 to 2 percent. Moving to our experimental results, we find that health inspectors with one standard deviation higher Big 5 index respond to the treatment by increasing health inspections by 27 percentage points in addition to the pure treatment effect of 10 percentage points. This is almost a 50 percent increase in inspections from the unconditional mean. Last, among senior health officials, we find that officials with high Big 5 index respond to data on absence of their subordinates much more aggressively. Doctors

in facilities overseen by senior health officials with a Big 5 index one standard deviation above the mean are 40 percentage points less likely to be absent in the month following a report that a facility is underperforming. This is an increase in doctor attendance of over 80 percent.

The paper proceeds as follows: Section 2 provides institutional details of the public health sector in Punjab, Pakistan, on which all of our analysis is focused. Section 3 outlines our research design, including the measures of policymaker personalities that were collected of doctors, health inspectors, and senior health officials, and a description of the policy experiment we will examine, "Monitoring the Monitors". Section 4 then outlines a simple model based on Almlund et al. (2011) to explain how personality traits can affect job task selection and performance. Section 5 then presents both non-experimental and experimental results on the association between personality traits and job performance. Section 6 concludes.

2 Background

2.1 The Public Health System

In Punjab province, the provision of health care services is managed by the Department of Health, which is based at the provincial headquarters in Lahore. There are five major types of facilities: (1) Basic Health Unit (BHU); (2) Rural Health Center (RHC); (3) Tehsil Headquarter Hospital[4] (THQ); (4) District Headquarter Hospital (DHQ); (5) Teaching Hospitals.

We focus on Basic Health Units (BHUs). BHUs are the smallest public health care units. They are designed to be the first stop for patients seeking medical treatment in government facilities. (Hereafter in this paper, we use the word 'clinic' interchangeably to describe BHUs). There are 2496 BHUs in Punjab.[5] They largely serve rural populations; almost all such clinics are exclusively operating in rural and peri-urban areas. These clinics provide

[4]In Punjab, a Tehsil is the largest sub-division of a district

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

several services, including out-patient services, neo-natal and reproductive healthcare, and vaccinations against diseases. 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.

2.1.1 Health Sector Administration

District governments are responsible for managing local health facilities. The District Health Department is headed by an Executive District Officer who reports both to the chief bureaucrat of the district and to the most senior provincial health officials.[6] He is supported by several Deputy District Officers, typically one for each tehsil.[7] Figure 1 depicts the (simplified) health administration hierarchy in Punjab, Pakistan.

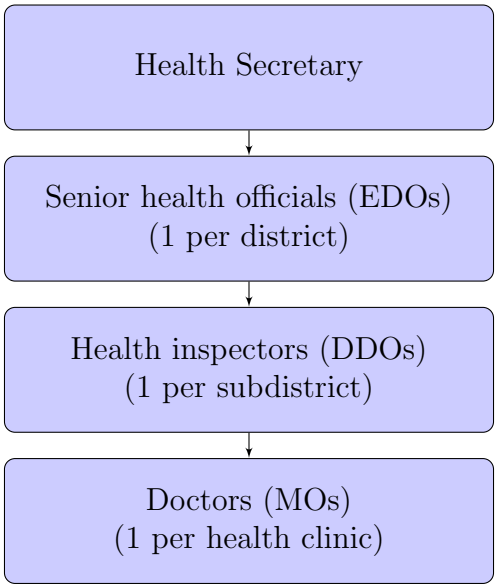
The central department has also established a parallel entity known as the Punjab Health Sector Reform Program (PHSRP). PHSRP is tasked with initiating programs to reform the primary health system with support from international and donor organizations. PHSRP is responsible for the implementation of the smartphone monitoring program we evaluate in this paper.

The Deputy District Officer is the lowest position in the officer-cadre of district health administration. He inspects all health facilities in a given Tehsil. This officer is required to visit every clinic at least once a month and record information collected during the visit on a standard form. The Deputy District Officer has authority to punish the clinic's absent staff by issuing a show-cause notice, suspension and withholding pay (in case of contract staff). The Executive District Officer relies entirely on this subordinate officer to ensure staff presence. As the administrative head of the health department in the district, the Executive District Officer desires smooth functioning of the setup at minimum acceptable

[6] The Director General of Health Services and the Secretary of the Health Department

[7] The Executive District Officer is also supported by other staff, but they are excluded for clarity because they are irrelevant to our discussion here.

Figure 1: Health sector administration in Punjab



level. He relies on the Deputy District Officer to ensure this smooth function by sanctioning underperforming facilities in terms of staff attendance, medicine availability and cleanliness etc. Throughout the paper, we will refer to Deputy District Officers as health inspectors and Executive District Officers as senior health officials, focusing on their role rather than their title.

3 Research Design and Experiment

Our research design links survey measures of personality with the performance of doctors in Punjab, both with correlations and experimentally in response to two related experimental policy interventions. In this section, we will first explain in detail our survey measures of personality. We will then outline the smart phone monitoring policy intervention that we implemented in Punjab as a randomized control trial, as well as the information salience experiment that we built into the broader intervention. For expositional ease, we will leave discussion of our performance measures to Section 5, when we discuss our results.

3.1 Measuring Personality

We interviewed 389 doctors, 101 health Deputy District Officers (DDOs), and 33 Executive District Officers (EDOs) across Punjab to put together our personality data. Doctors were interviewed at their BHUs during the second and third waves of our independent inspections, described below, as well as during a special round of follow-ups conducted in November 2012. Interviews of DDOs and EDOs were conducted in November and December 2012.[8]

Our partnership with PHSRP meant that doctors, DDOs, and EDOs were directly instructed to respond to our surveys, subject to their comfort with the questions being asked and all other usual human subjects requirements being honored. For DDOs and EDOs, this resulted in our surveying the entire population within Punjab in one round, subject to changes in staffing during the course of our experiment (which were very few given the short timeframe). Doctors, however, are transferred from one clinic to another more often, are absent from their workplace a majority of the time, and our posted at rural clinics that are generally harder to visit than district headquarters. This led to our three different attempts to interview doctors. The first two were unannounced as part of our independent inspections. To maximize the effectiveness of our third attempt, we obtained the phone numbers of all posted doctors from PHSRP, and our enumerators called ahead and scheduled survey meetings with each doctor that had not been present during the first two attempts. In the end, this resulted in our surveying 389 of roughly 544 posted doctors, or 72 percent of our sample population.

The measures of personality in this paper were drawn from rich and growing literatures within psychology and economics. The first measure of personality, the Big 5 personality index, was first developed by psychologists in the 1980s and has subsequently become the standard and most widely used personality taxonomy in the field. [9] Note that, to our knowl-

[8]Note that we consider our sample of DDOs and EDOs to be the universe. Though these samples do not represent those overseeing all sub-districts or districts respectively, they represent all of those DDOs and EDOs that were appointed at the time of our surveys. There are frequently vacancies in these positions.

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

edge, only one paper has used these measures to better understand public sector employees in the developing context. In Dal Bo et al. (2013), the authors find that offering more competitive wages to government employees causes more desirable employees to apply for jobs as measured by the same personality measures. We chose our personality measures in part to follow this paper.

The Big 5 personality index consists of five traits—openness, conscientiousness, extraversion, agreeableness, and neuroticism. We measured these traits using a 60 question survey developed specifically in Urdu and validated for use in Pakistan. Each trait is measured separately as the sum of 12 questions, and all traits were normalized into z-scores and averaged to form a single Big 5 index. Each question 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 5-point Likert scale (Disagree strongly, Disagree a little, Neutral, Agree a little, or Agree strongly). See John et al. (2008) for details on each trait and the questions used.

We measure each health official's public service motivation using the Perry Public Service Motivation Index, also a very widely used measure in psychology, developed around the same time period as the Big 5 index and similarly validated. It takes the same form as the Big 5 questionnaire, with 40 total questions and six specific dimensions—attraction to policymaking, commitment to policymaking, social justice, civic duty, compassion, and self-sacrifice.[10]

We also measured prosocial behavior using a series of questions on charity work, volunteer work, voting history, religious activity, etc. We also measured time use, and ask a rich set of questions on political motivation, which we will explore further in the future.

We include a copy of our personality survey in the appendix. Though the survey included is for doctors (medical officers), we used the exact same instrument for health inspectors and senior health officials. We include both the formatted, Urdu version that was fielded as well

[10]See Perry and Wise (1990) and Perry (1996) for the development of the measure, and Petrovsky (2009) for a synthesis of the empirical research using this measure.

as a translation of the instrument to English for reference.

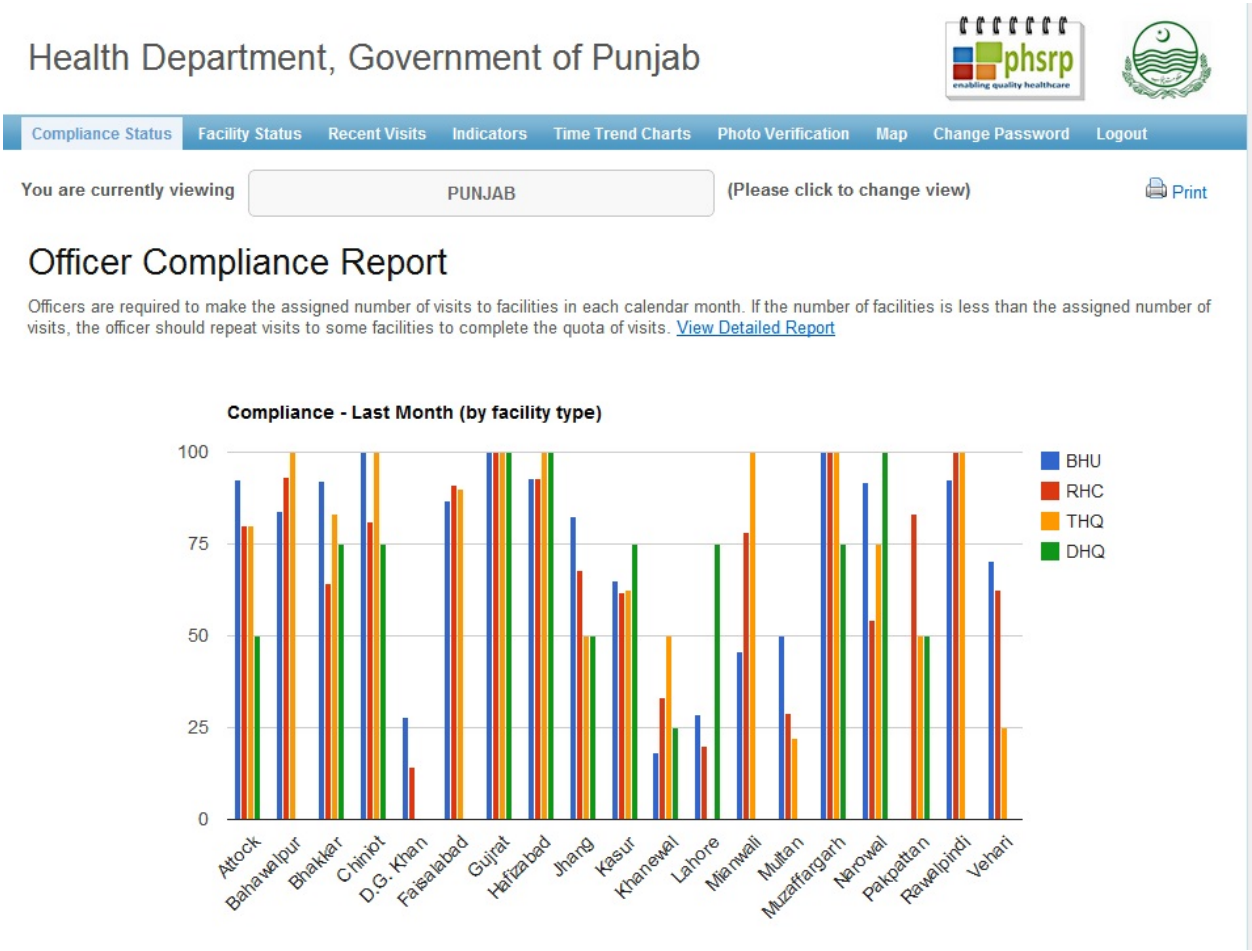
3.2 Monitoring the Monitors

We collected personality data during a larger experimental policy reform aimed at exploring the use of audits by government monitors as a solution to the problem of absence. The "Monitoring the Monitors" program replaced the traditional paper-based monitoring system, which collects data on facility utilization, resource availability, and worker absence, with an android-based smartphone application. Data are transmitted to a central database using a General Packet Radio Service (GPRS) in real time. Data are then aggregated and summary statistics, charts, and graphs are presented in a format designed in collaboration with senior health officials. That data are: (i) aggregated in the province in real time; (ii) geo-tagged, time-stamped, and complemented with facility staff photos to check for reliability; and (iii) available in real time to district and provincial officers through an online dashboard. Figure 2 shows one view of the online dashboard.

In addition to aggregating information, the online dashboard provides for a second experiment. In this experiment, we manipulate the salience of the information provided through the online dashboard to senior health officials about the performance of doctors and other health workers in their districts. We make certain facilities salient by highlighting clinic inspection reports that find three or more staff to be absent in red. This highlighting is the only difference between any two entries on the website. Thus as the cutoff between two or less and three or more staff was arbitrarily determined by our team and never communicated to senior health officials as important in any other way, we can invoke a regression discontinuity (RD) framework for understanding the impact of information salience to a specific set of decision makers after a policy change. See Section 5 for more details on the discontinuity and for a discussion of our empirical methodology given this RD research design.

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

Figure 2: Online Dashboard - Summary of Inspection Compliance by District



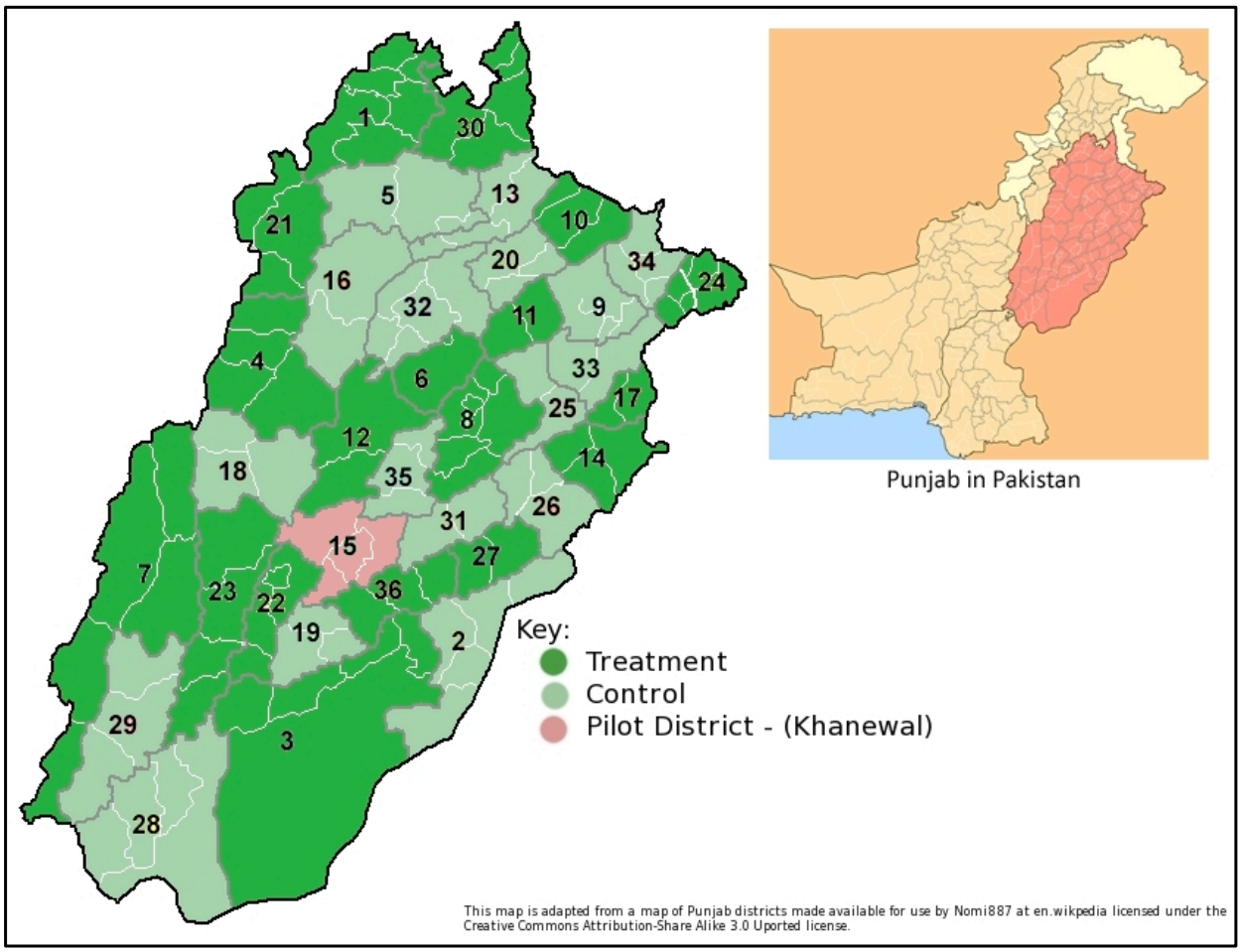
of Khanewal. We remove Khanewal district from the experimental sample. Health administration staff were provided with smartphones and trained to use the application. The main purpose of the pilot was to ensure that the technology was working and to refine the application and the dashboard. During the pilot, several inspectors requested that the program require pictures of all staff in attendance, not just the inspector because they thought it might reduce pressure from health staff to falsify attendance.

Our experimental sample comprised 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 stood to benefit from the program. While we have administrative data for all facilities, we monitor a subsample of 850 clinics, drawn to be representative of facilities in the province, using independent inspections. 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 3 depicts control and treatment districts.

We randomized at the district level. The intervention channels information about inspections to district health officials; randomization at a finer level is therefore very likely to generate externalities. The Department of Health also determined that sub-district randomization was 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 much weaker across districts.

See Callen et al. (2013) for the core results from the broad Monitoring the Monitors experiment as well as for the information salience experiment contained within it. In this paper, we take these experiments as more general policy interventions, and seek to understand how doctor, inspector, and executive district officer personality affects response to the reform.

Figure 3: Treatment and Control Districts



4 Model

At its core, the policy intervention that we analyze in this paper can be seen as an increase in the probability that health inspector and doctor shirking is detected. We examine whether doctors and health inspectors with comparatively better personality traits are more or less sensitive to this increase in the detection probability. Intuitively, a fixed increase in the detection probability might induce better types to work, as working is less costly. On the other hand, better types might have better outside options, and so be less sensitive to increases in the probability of detection. The following framework provides a simple description of the circumstances under which high types are more likely to be induced to work by increased monitoring.

Let us consider the binary decision that a doctor or inspector makes of whether to show up to work or to shirk. Let the cumulative distribution of worker types be $F(\theta)$. 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. The marginal worker indifferent between working and shirking and will satisfy

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

If we assume that $\frac{\partial \lambda}{\partial \theta} < 0$, then it is straightforward to see that all workers with θ greater than that of the marginal worker will choose to work. Intuitively, this means it costs better type individuals less to exert effort at work. This is inline 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. This might be because better type workers are more efficient with their time, or because their psychic costs are lower given that they are more motivated individuals, etc.

In order to understand the effect of a policy intervention on the decision to work, let us

solve for the marginal worker type, or θ^M

$$\theta^M = \lambda^{-1}(p(W + c)) \tag{2}$$

Now we can see how this marginal worker type changes with an increase in detection probability:

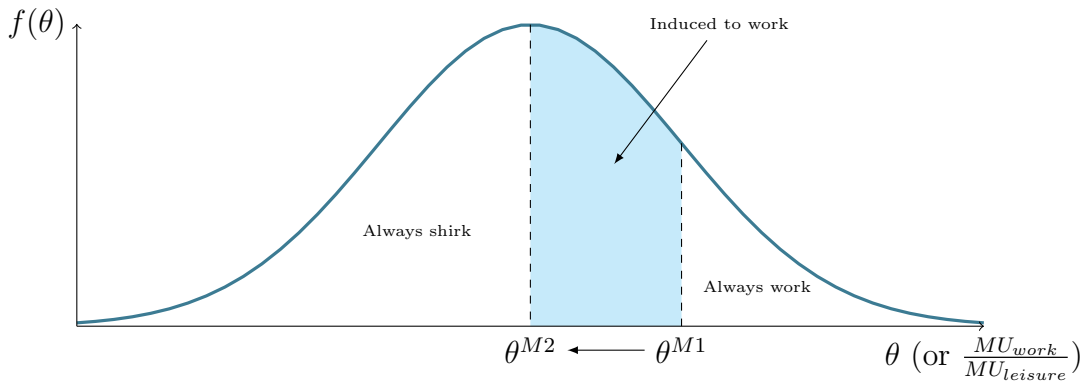
$$\frac{\partial \theta^M}{\partial p} = \frac{1}{\lambda'(\lambda^{-1}(p(W + c))} \tag{3}$$

Given our earlier assumption that $\frac{\partial \lambda}{\partial \theta} < 0$, and assuming that $p, W, c > 0$, it must be that $\frac{\partial \theta^M}{\partial p} < 0$, or that the marginal worker decreases with an increase in detection probability. We can make this point more intuitively by considering equation (1). If p increases, then for $p, W, C > 0$, we see that the right hand side of the equation will become unambiguously smaller. To solve this updated indifference equation for θ^M, the left hand side must also decrease from before the increase in p, and this can only occur through a decrease in θ^M.

Thus in this simple case the marginal worker type decreases as the detection probability increases. And, since it remains that all workers with θ greater than that of the new marginal worker will choose to work, an increase in p will lead to an increase in workers showing up at work, and the new workers will be those with values of θ that lie between the original marginal θ and the new, lower marginal θ. This means that the first workers to shift from shirking to work with an increase in detection probability are those with the best types of those that shirk beforehand.

We can see this in a simple picture in figure 4. Let θ^{M1} be the marginal worker before an increase in p and θ^{M2} the lower-type marginal worker afterwards. The shaded area thus represents those types that are induced to work given this increase in the probability of detection.

Figure 4: Effect of an increase in detection probability on the decision to work or shirk



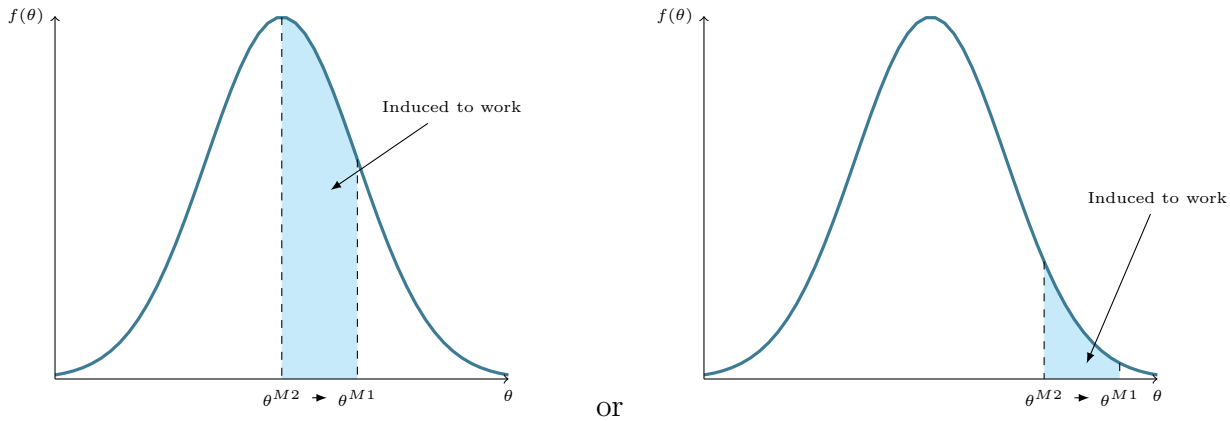
Two of the predictions from this model coincide with the two primary relationships that we test in this paper. First, we predict that prior to any intervention, those workers with better personalities will have better measured performance (assuming $0 < \theta^M < 1$ in the case when θ is normalized to be between 0 and 1). The second is that an increase in the detection probability of doctors and inspectors shirking will specifically induce those workers with better personalities to work.[11]

Figure 5 makes two additional points. The first is perhaps obvious—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. We should be careful not to attribute our effects to these individuals. The second point is that the initial position of θ^M potentially matters significantly to the size of the impact of an increase in detection probability. In the left panel, a large proportion of the population of workers is induced to work from an increase in p. In the right panel, an increase in p leading to the same level change in θ^M leads to a much smaller proportion of workers being induced to work. This also highlights the importance of the shape of the distribution of types, as a very narrow distribution might see very different effects than a uniform distribution from an increase in p, for example. This increases the policy relevance of the results in this paper. If

[11]Note that we have an entire paper dedicated to the obvious third prediction of this model—that an increase in detection probability will induce workers to work more overall. See Callen et al. (2013).

we believe that personality measures serve as a proxy for type, both the initial position of θ^M and the shape of the distribution of types can be estimated ex-ante using these personality measures, allowing for better targeted policies.

Figure 5: Understanding the 'local' in estimated LATEs

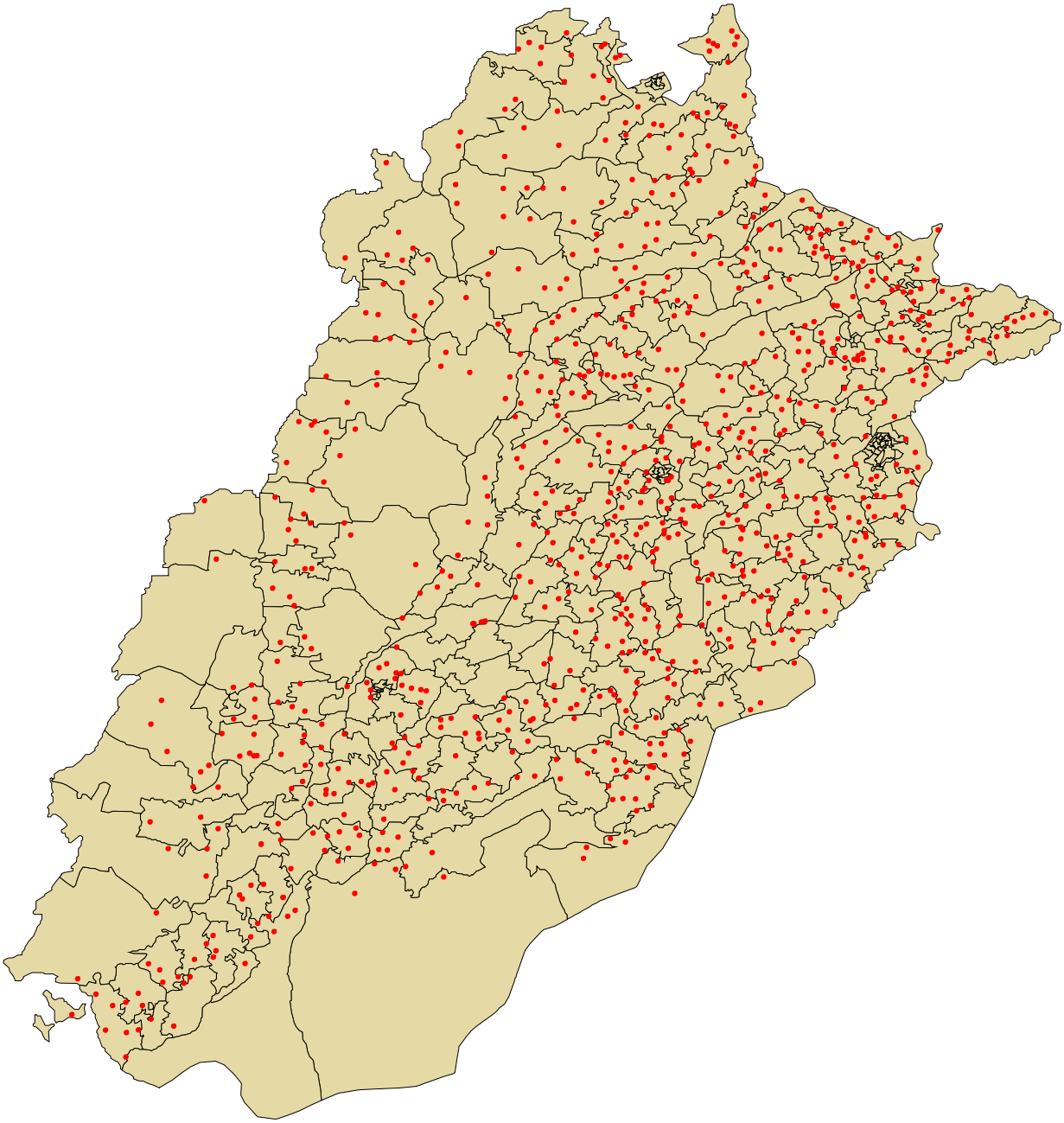


5 Data and Results

5.1 Does Personality Predict Performance?

For our outcome measures, we collected primary data on a representative sample of 850 of the 2,496 Basic Health Units in Punjab (34 percent). BHUs were selected randomly using an Equal Probability of Selection (EPS) design, stratified on district and distance between the district headquarters and the BHU. Therefore, our estimates of absence are self-weighting, and so no sampling corrections are used in the analysis. All districts in Punjab except Khanewal are represented in our data. To our knowledge, this is the first representative survey of BHUs in Punjab. Figure 6 provides a map of the Basic Health Units in our experimental sample along with the different Provincial Assembly constituencies in Punjab.

Figure 6: Locations of Basic Health Units in the Experimental Sample



In our sample of 850 clinics, we collected data through independent inspection. We made unannounced visits to these facilities three times, first in November 2011, then in June 2012, and finally in October 2012. We collected information on staff absence, health inspections, and facility usage. Our survey team interviewed the Medical Officer, or doctor, the Dispenser or Health/Medical Technician, and the Lady Health Visitor before physically verifying the attendance of the Mid-Wife and the School Health and Nutrition Specialist. These 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 BHUs 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. Data collection and entry followed back-checks and other validation processes consistent with academic best practice.

Table 1 gives the personality summary statistics of the doctors in our survey. The main take-aways from these tables is that there is a large amount of heterogeneity in personality type for these health officials. This is not a surprise given past use of these measures, and given the fact that the Big5 index, for example, was created in a way as to maximize explanatory variation.[12]

Table 2 shows the summary stats of the universe of health inspectors in control districts of Punjab. We can see that their personality traits are very similar to surveyed doctors. This allows us to make the broader point that these measures can allow for comparisons of individuals across occupation. Of course, most health inspectors in our sample are doctors themselves by training, but they serve a much different role in the bureaucracy, are likely to be on a different career path, and likely face different political pressures than the doctors in our sample. [13]

We correlate personality measures for doctors with two measures of job performance: (i)

[12]See Borghans et al. (2008) for an explanation of the Big5 index development process.

[13]Perhaps a stronger argument in favor of this point is the fact that the trait distributions from our sample are very similar to those from Dal Bo et al. (2013), which surveys government promoters in Mexico.

Table 1: Doctor personality summary statistics (control districts)

	Mean	SD	P10	P50	P90	Obs
Personality traits						
Big 5 index (z-score)	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	4.00	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 (z-score)	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	5.00	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	4.00	4.60	192
Prosocial behavior						
Prosocial (principal comp.)	0.03	1.26	-1.88	0.43	1.42	184
Did charity work in the past year (=1)	0.90	0.30	1.00	1.00	1.00	192
Has ever run for office (=1)	0.03	0.16	0.00	0.00	0.00	192
Has ever volunteered (=1)	0.86	0.34	0.00	1.00	1.00	191
Voted in the last NA election (=1)	0.55	0.50	0.00	1.00	1.00	192
Has ever donated blood (=1)	0.77	0.42	0.00	1.00	1.00	192
Goes to the mosque regularly (=1)	0.48	0.50	0.00	0.00	1.00	184
Believes people can be trusted (=1)	0.55	0.50	0.00	1.00	1.00	192
Performance						
Present (=1)	0.23	0.42	0.00	0.00	1.00	1197

Notes: Sample: doctors in control districts that completed the personalities survey module, given in waves 2 and 3 and during a special follow-up round. Doctors were only asked to complete the module once. All personality traits and public sector motivation variables measured on a one to five Likert scale unless otherwise indicated. Performance sample is clinic-wave observations in control districts across waves 1 through 3.

Table 2: Health inspector personality summary statistics (control districts)

	Mean	SD	P10	P50	P90	Obs
Personality traits						
Big 5 index (z-score)	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 (z-score)	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	5.00	49
Commitment	3.97	0.37	3.43	3.86	4.50	49
Compassion	3.66	0.49	3.00	3.62	4.25	49
Self Sacrifice	4.40	0.45	3.86	4.50	5.00	49
Social justice	4.20	0.43	3.60	4.20	5.00	49
Prosocial behavior						
Prosocial (principal comp.)	0.19	1.05	-1.06	0.58	1.42	49
Did charity work in the past year (=1)	0.94	0.24	1.00	1.00	1.00	49
Has ever run for office (=1)	0.02	0.14	0.00	0.00	0.00	49
Has ever volunteered (=1)	0.92	0.28	1.00	1.00	1.00	49
Voted in the last NA election (=1)	0.71	0.46	0.00	1.00	1.00	49
Has ever donated blood (=1)	0.80	0.41	0.00	1.00	1.00	49
Goes to the mosque regularly (=1)	0.37	0.49	0.00	0.00	1.00	49
Believes people can be trusted (=1)	0.59	0.50	0.00	1.00	1.00	49
Performance						
DDO inspected in the last two months (=1)	0.53	0.50	0.00	1.00	1.00	1263

Notes: Sample: Health inspectors in control districts that completed the personalities survey module. Performance sample is clinic-wave observations in control districts across waves 1 through 3.

whether doctors were present during our unannounced visits, and (ii) a proxy measure of collusion between doctors and health inspectors to falsify inspection reports. Similarly, we correlate personality measures for health inspectors with two measures of job performance: (i) whether health inspectors inspected each health clinic in the month prior to an unannounced visit (they are supposed to inspect each facility each month), and (ii) the same proxy measure of collusion between doctors and health inspectors.

In the correlations for both doctors and health inspectors, we define collusion as a dummy variable coded as one when a doctor is reported absent in both of the unannounced visits in our survey waves two and three but is reported as present by health inspectors during every visit between the launch of the program and present (up to 73 visits). The type of collusion that we have in mind here is when a health inspector calls up a doctor the day before an inspection to alert him or her to be in attendance. Then, after the health inspector records his or her presence, the doctor is under very little pressure to attend until he or she get another phone call the following week or month. 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.[14]

We can see from figure 7, Panel A, that doctors that score one standard deviation higher on the Big 5 measure of conscientiousness are about 5 percent more likely to be present at work during an unannounced visit. In absolute terms, doctors with one point higher on the Likert scale for conscientiousness are 9.8 percentage points more likely to be present.[15] Extroversion is similarly predictive, as are two measures of PSM—civic duty and self sacrifice. And in general, the direction of all but one of the other coefficients is the same.

We can see in Panel B that doctor personality measures are even stronger predictors of potential collusion between health inspectors and doctors than they are of attendance. Ten of 11 Big5 and PSM traits are highly predictive of collusion and with the signs we

[14]Results provided upon request.

[15]See appendix table A.1 for non-standardized point-estimates

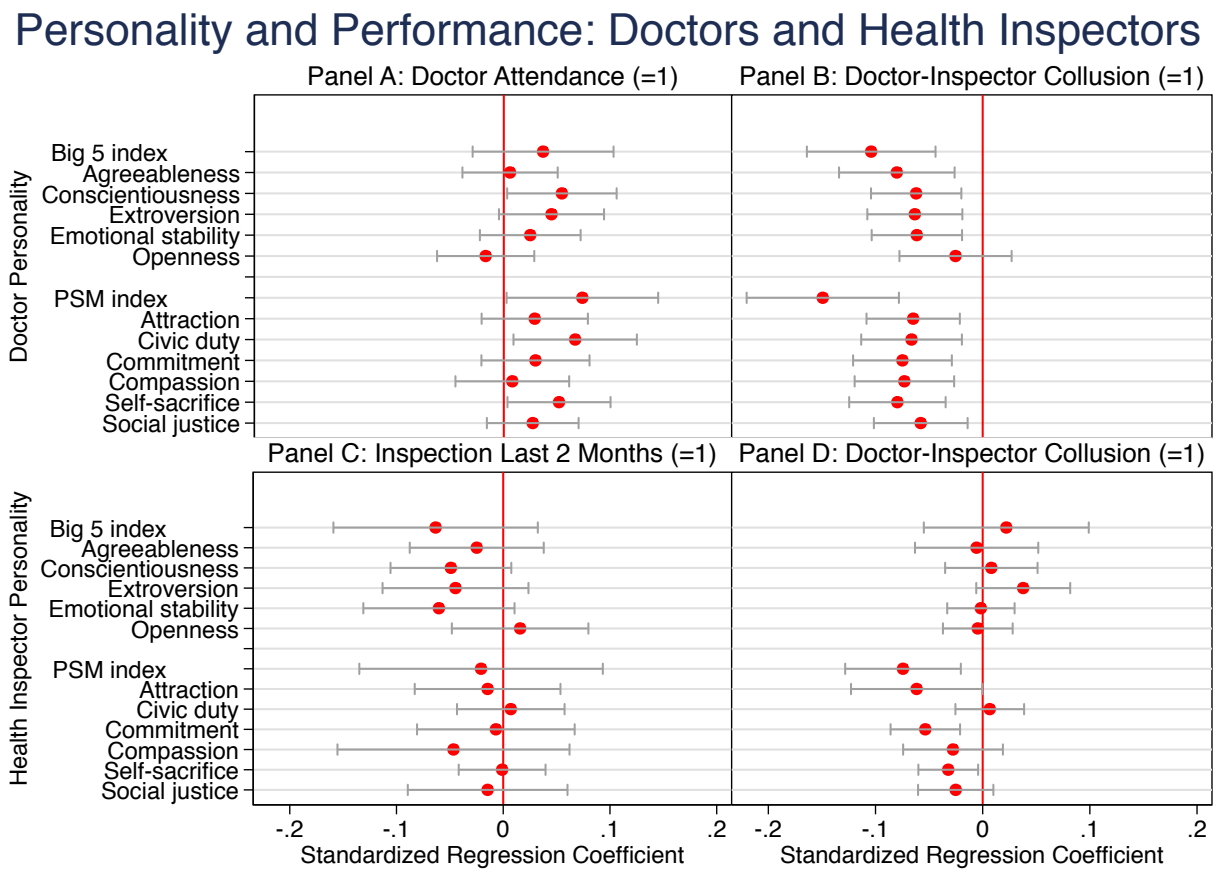


Figure 7: Each regression coefficient reported comes from a separate regression of the displayed performance measure on the displayed doctor or health inspector personality measure. Error bars represent 95 percent confidence intervals, with standard errors clustered at the clinic levels. All regressions 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. Sample: control district clinics for which doctor personality data is available and a doctor is posted.

would expect. Doctors who score one standard deviation higher on measured civic duty, for example, are about 6 percentage points likely to be identified as potentially colluding. In absolute terms, doctors who score one point higher on the Likert scale for civic duty are 11.4 percentage points less likely to be identified as potentially colluding.[16]

The correlations between health inspector personality measures and performance are not so consistent. In Panel C, all coefficients are not significantly different from zero. And, in 11 of 13 cases the coefficient has the opposite sign as in the case of doctors. In Panel D, we see that PSM traits are associated with less collusion, enough to make the aggregate z-score, attraction, commitment, and self-sacrifice statistically different from zero. In this case, health inspectors that score one standard deviation higher on aggregate PSM are about 7 percentage points less likely to be identified as potentially colluding.[17] There can be many explanations for why health inspector personality measures are less strong predictors of performance. One thing to note is that the inspections performance measure is different in scope from the attendance measure for doctors. This is because inspections are only one of the tasks that health inspectors preform. They also have office duties that would allow them to show up to work but not conduct inspections.

Though non-experimental, we take figure 7 as a rather strong validation of our personality measures in predicting the performance of public sector employees in Pakistan. Together, we find that increases in 20 of the 52 personality traits we examine are associated with increased job performance for doctors and health inspectors at 95 percent confidence. And in none of these cases does is a personality measure significantly associated with decreased job performance. And we take the fact that the health inspector results are not as strong as those for doctors as a reminder of the importance of understanding personality not just of supervised employees but also of inspectors, such as DDOs, and other supervisors, such as senior health officials (EDOs).

[16]See appendix table A.2 for non-standardized point-estimates

[17]See appendix tables A.3 and A.4 for non-standardized point-estimates.

5.2 Heterogeneous Monitoring the Monitors Effects

We now move on to our experimental analysis of how personality impacts the response to doctors to the Monitoring the Monitors policy reform explained above. As such, our core identification strategy is that of heterogeneous treatment effects. We interact our measures of personality with experimentally valid treatment assignment ('Monitoring'). Assuming that personality as measured does not change as a result of the program, we will be able to interpret these results causally.

Tables 3 and 4 present the treatment balance table for our personality measures for doctors and health inspectors. There are no significant differences in the treatment and control districts between mean personalities of doctors.

Tables 5 and 6 presents our first set of experimental results. As we can see, with the exception of social justice, we find no significant heterogeneous effects of our experiment on doctor attendance. This is not surprising given the fact that we find no overall positive effect on doctor attendance of the treatment (the first column in each panel). This is, however, surprising given how predictive personality is of doctor performance in control districts. This difference could be because of OVB in the case of our correlations, or more likely it is because there is something different about showing up to work in the first place and responding to a treatment. We are currently evaluating this difference at length.

With health inspectors, on the other hand, there are very strong heterogeneous effects of our experiment on the rate of health inspections. A one standard deviation higher Big 5 index for health inspectors, for example, is associated with a 27 percentage point increase in health inspections in the two months prior to one of our unannounced visits at a clinic. On an unconditional mean of 58 percent, this is a 47 percent increase in the rate of inspections. This is clearly an economically significant magnitude. This effect is being driven most strongly by emotional stability, which is a commonly important trait within the literature in economics examining the Big 5 measure, but note that all Big 5 traits have positive and large coefficients. We also see some positive and similarly large effects of attraction, compassion,

Table 3: Treatment balance on doctor personality

	Treatment	Control	Difference	P-value
Big 5 personality traits				
Big 5 index (z-score)	-0.058	0.042	-0.100	0.295
	[0.713]	[0.820]	[0.095]	.
Agreeableness	3.498	3.577	-0.079	0.309
	[0.622]	[0.678]	[0.077]	.
Conscientiousness	3.958	3.996	-0.037	0.605
	[0.548]	[0.570]	[0.072]	.
Extroversion	3.624	3.686	-0.062	0.277
	[0.464]	[0.501]	[0.057]	.
Emotional stability	-2.647	-2.536	-0.111	0.180
	[0.641]	[0.702]	[0.082]	.
Openness	2.926	2.932	-0.006	0.907
	[0.372]	[0.451]	[0.050]	.
Public service motivation				
PSM index (z-score)	-0.017	-0.018	0.001	0.989
	[0.695]	[0.691]	[0.079]	.
Attraction	3.481	3.442	0.039	0.581
	[0.630]	[0.610]	[0.070]	.
Civic duty	4.182	4.184	-0.002	0.969
	[0.594]	[0.526]	[0.059]	.
Commitment	3.773	3.774	-0.001	0.982
	[0.511]	[0.463]	[0.050]	.
Compassion	3.493	3.546	-0.053	0.432
	[0.515]	[0.516]	[0.067]	.
Self Sacrifice	4.065	4.080	-0.015	0.820
	[0.563]	[0.574]	[0.065]	.
Social justice	3.950	3.906	0.044	0.464
	[0.571]	[0.619]	[0.060]	.
Prosocial behavior				
Prosocial (principal comp.)	-0.121	0.020	-0.141	0.305
	[1.219]	[1.283]	[0.137]	.
Did charity work in the past year (=1)	0.872	0.905	-0.032	0.352
	[0.333]	[0.295]	[0.035]	.
Has ever run for office (=1)	0.029	0.027	0.002	0.932
	[0.168]	[0.163]	[0.021]	.
Has ever volunteered (=1)	0.866	0.857	0.009	0.814
	[0.340]	[0.351]	[0.036]	.
Voted in the last NA election (=1)	0.587	0.558	0.029	0.622
	[0.493]	[0.498]	[0.059]	.
Has ever donated blood (=1)	0.700	0.752	-0.051	0.309
	[0.458]	[0.432]	[0.050]	.
Goes to the mosque regularly (=1)	0.434	0.457	-0.023	0.685
	[0.496]	[0.500]	[0.057]	.
Believes people can be trusted (=1)	0.504	0.571	-0.067	0.233
	[0.500]	[0.497]	[0.056]	.
# Observations	242	147		

Notes: Variable standard deviations reported in brackets. Standard errors clustered at the district level reported in parentheses. Actual observations for each regression vary by a small amount based on no responses. Sample limited to clinics where a doctor is posted.

Table 4: Treatment balance on inspector personality

	Treatment	Control	Difference	P-value
Big 5 personality traits				
Big 5 index (z-score)	-0.017	0.018	-0.035	0.802
	[0.637]	[0.745]	[0.140]	.
Agreeableness	3.783	3.659	0.124	0.231
	[0.477]	[0.541]	[0.103]	.
Conscientiousness	4.159	4.117	0.041	0.679
	[0.452]	[0.536]	[0.100]	.
Extroversion	3.703	3.734	-0.031	0.754
	[0.525]	[0.459]	[0.099]	.
Emotional stability	-2.461	-2.338	-0.124	0.307
	[0.571]	[0.624]	[0.120]	.
Openness	3.020	3.113	-0.093	0.264
	[0.471]	[0.350]	[0.083]	.
Public service motivation				
PSM index (z-score)	-0.061	0.071	-0.131	0.288
	[0.621]	[0.614]	[0.123]	.
Attraction	3.552	3.568	-0.016	0.881
	[0.532]	[0.568]	[0.110]	.
Civic duty	4.255	4.435	-0.180	0.034
	[0.415]	[0.424]	[0.084]	.
Commitment	3.915	3.969	-0.054	0.514
	[0.458]	[0.370]	[0.083]	.
Compassion	3.743	3.659	0.085	0.380
	[0.475]	[0.488]	[0.096]	.
Self Sacrifice	4.316	4.395	-0.079	0.396
	[0.482]	[0.454]	[0.093]	.
Social justice	4.098	4.200	-0.102	0.268
	[0.490]	[0.430]	[0.092]	.
Prosocial behavior				
Prosocial (principal comp.)	0.059	0.195	-0.136	0.551
	[1.217]	[1.050]	[0.227]	.
Did charity work in the past year (=1)	0.922	0.939	-0.017	0.739
	[0.272]	[0.242]	[0.051]	.
Has ever run for office (=1)	0.060	0.020	0.040	0.320
	[0.240]	[0.143]	[0.040]	.
Has ever volunteered (=1)	0.863	0.918	-0.056	0.377
	[0.348]	[0.277]	[0.063]	.
Voted in the last NA election (=1)	0.745	0.714	0.031	0.732
	[0.440]	[0.456]	[0.090]	.
Has ever donated blood (=1)	0.784	0.796	-0.012	0.888
	[0.415]	[0.407]	[0.082]	.
Goes to the mosque regularly (=1)	0.431	0.367	0.064	0.518
	[0.500]	[0.487]	[0.099]	.
Believes people can be trusted (=1)	0.580	0.592	-0.012	0.906
	[0.499]	[0.497]	[0.100]	.
# Observations	52	49		

Notes: Variable standard deviations reported in brackets. Standard errors clustered at the district level reported in parentheses. Actual observations for each regression vary by a small amount based on no responses.

Table 5: Personalities and doctor attendance

	Doctor attendance (=1)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: Personality traits									
Monitoring (=1)			0.019	0.022	0.144	-0.232	-0.073	-0.061	-0.006
			(0.076)	(0.077)	(0.417)	(0.495)	(0.374)	(0.252)	(0.530)
Monitoring x Big5 index				0.005					
				(0.086)					
Monitoring x Agreeableness					-0.033				
					(0.116)				
Monitoring x Conscientiousness						0.063			
						(0.123)			
Monitoring x Extroversion							0.026		
							(0.097)		
Monitoring x Emotional stability								-0.031	
								(0.090)	
Monitoring x Openness									0.009
									(0.177)
Mean of dependent variable			0.540	0.540	0.540	0.540	0.540	0.540	0.540
# Districts			34	34	34	34	34	34	34
# Clinics			474	474	474	474	474	474	474
# Observations			1216	1216	1216	1216	1216	1216	1216
R-Squared			0.013	0.013	0.016	0.013	0.013	0.013	0.013
PANEL B: Public service motivation									
Monitoring (=1)		0.019	0.020	-0.123	-0.067	0.231	0.296	-0.114	-1.058***
		(0.076)	(0.076)	(0.313)	(0.547)	(0.409)	(0.369)	(0.535)	(0.327)
Monitoring x PSM index			0.057						
			(0.086)						
Monitoring x Attraction				0.040					
				(0.085)					
Monitoring x Civic duty					0.021				
					(0.125)				
Monitoring x Commitment						-0.056			
						(0.111)			
Monitoring x Compassion							-0.077		
							(0.106)		
Monitoring x Self sacrifice								0.033	
								(0.135)	
Monitoring x Social justice									0.273***
									(0.090)
Mean of dependent variable		0.540	0.540	0.540	0.540	0.540	0.540	0.540	0.540
# Districts		34	34	34	34	34	34	34	34
# Clinics		474	474	474	474	474	474	474	474
# Observations		1216	1216	1216	1216	1216	1216	1216	1216
R-Squared		0.013	0.018	0.016	0.013	0.019	0.016	0.013	0.027
PANEL C: Prosocial behavior									
Monitoring (=1)	0.027	0.029	0.042	0.017	0.040	0.027	-0.019	0.081	0.076
	(0.078)	(0.077)	(0.209)	(0.077)	(0.164)	(0.092)	(0.145)	(0.102)	(0.099)
Monitoring x Prosocial (principal comp.)		0.001							
		(0.050)							
Monitoring x Did charity work in the past year (=1)			-0.028						
			(0.197)						
Monitoring x Has ever run for office (=1)				0.069					
				(0.335)					
Monitoring x Has ever volunteered (=1)					-0.023				
					(0.157)				
Monitoring x Voted in the last NA election (=1)						-0.006			
						(0.108)			
Monitoring x Has ever donated blood (=1)							0.061		
							(0.163)		
Monitoring x Goes to the mosque regularly (=1)								-0.109	
								(0.138)	
Monitoring x Believes people can be trusted (=1)									-0.108
									(0.131)
Mean of dependent variable	0.539	0.539	0.542	0.539	0.540	0.540	0.540	0.538	0.541
# Districts	33	33	34	34	34	34	34	33	34
# Clinics	453	453	470	472	472	473	473	457	472
# Observations	1156	1156	1206	1211	1212	1214	1214	1167	1211
R-Squared	0.010	0.011	0.012	0.012	0.014	0.017	0.018	0.014	0.013

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the clinic level reported in parentheses. All regressions include Tehsil (subdistrict) and survey wave fixed effects. Sample: control district clinics for which doctor personality data is available.

Table 6: Personalities and health inspections

	Health inspection in last two months (=1)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: Personality traits									
Monitoring (=1)			0.111	0.101	-0.671	-1.107	-0.311	0.815**	-1.022
			(0.115)	(0.101)	(0.646)	(0.794)	(0.633)	(0.324)	(0.692)
Monitoring x Big5 index				0.271*					
				(0.135)					
Monitoring x Agreeableness					0.215				
					(0.167)				
Monitoring x Conscientiousness						0.295			
						(0.185)			
Monitoring x Extroversion							0.114		
							(0.162)		
Monitoring x Emotional stability								0.305**	
								(0.128)	
Monitoring x Openness									0.370
									(0.228)
Mean of dependent variable			0.575	0.575	0.575	0.575	0.575	0.575	0.575
# Districts			35	35	35	35	35	35	35
# Clinics			707	707	707	707	707	707	707
# Observations			2115	2115	2115	2115	2115	2115	2115
R-Squared			0.062	0.082	0.085	0.080	0.064	0.081	0.073
PANEL B: Public service motivation									
Monitoring (=1)		0.121	0.110	-1.022**	0.648	-0.282	-0.530	-0.122	-0.752
		(0.112)	(0.105)	(0.473)	(0.682)	(0.688)	(0.784)	(0.884)	(0.713)
Monitoring x PSM index			0.160						
			(0.140)						
Monitoring x Attraction				0.316**					
				(0.123)					
Monitoring x Civic duty					-0.124				
					(0.154)				
Monitoring x Commitment						0.098			
						(0.165)			
Monitoring x Compassion							0.175		
							(0.199)		
Monitoring x Self sacrifice								0.056	
								(0.189)	
Monitoring x Social justice									0.206
									(0.163)
Mean of dependent variable		0.567	0.567	0.567	0.567	0.567	0.567	0.567	0.567
# Districts		35	35	35	35	35	35	35	35
# Clinics		721	721	721	721	721	721	721	721
# Observations		2157	2157	2157	2157	2157	2157	2157	2157
R-Squared		0.063	0.072	0.079	0.065	0.077	0.066	0.063	0.073
PANEL C: Prosocial behavior									
Monitoring (=1)	0.117	0.107	0.136	0.128	0.036	0.154	0.268	0.122	0.006
	(0.113)	(0.107)	(0.200)	(0.114)	(0.191)	(0.192)	(0.202)	(0.118)	(0.168)
Monitoring x Prosocial (principal comp.)		0.022							
		(0.073)							
Monitoring x Did charity work in the past year (=1)			-0.026						
			(0.222)						
Monitoring x Has ever run for office (=1)				-0.337***					
				(0.116)					
Monitoring x Has ever volunteered (=1)					0.087				
					(0.201)				
Monitoring x Voted in the last NA election (=1)						-0.048			
						(0.200)			
Monitoring x Has ever donated blood (=1)							-0.198		
							(0.195)		
Monitoring x Goes to the mosque regularly (=1)								0.004	
								(0.189)	
Monitoring x Believes people can be trusted (=1)									0.187
									(0.175)
Mean of dependent variable	0.566	0.566	0.566	0.565	0.566	0.566	0.566	0.566	0.565
# Districts	35	35	35	35	35	35	35	35	35
# Clinics	719	719	719	703	719	719	719	719	703
# Observations	2151	2151	2151	2103	2151	2151	2151	2151	2103
R-Squared	0.061	0.066	0.063	0.071	0.063	0.065	0.064	0.074	0.065

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard errors clustered at the clinic level reported in parentheses. All regressions include Tehsil (subdistrict) and survey wave fixed effects. Sample: control district clinics for which health inspector personality data is available.

and social justice within the PSM traits, the only attraction is significant. [18]

Figure 8: Nonparametric treatment effect

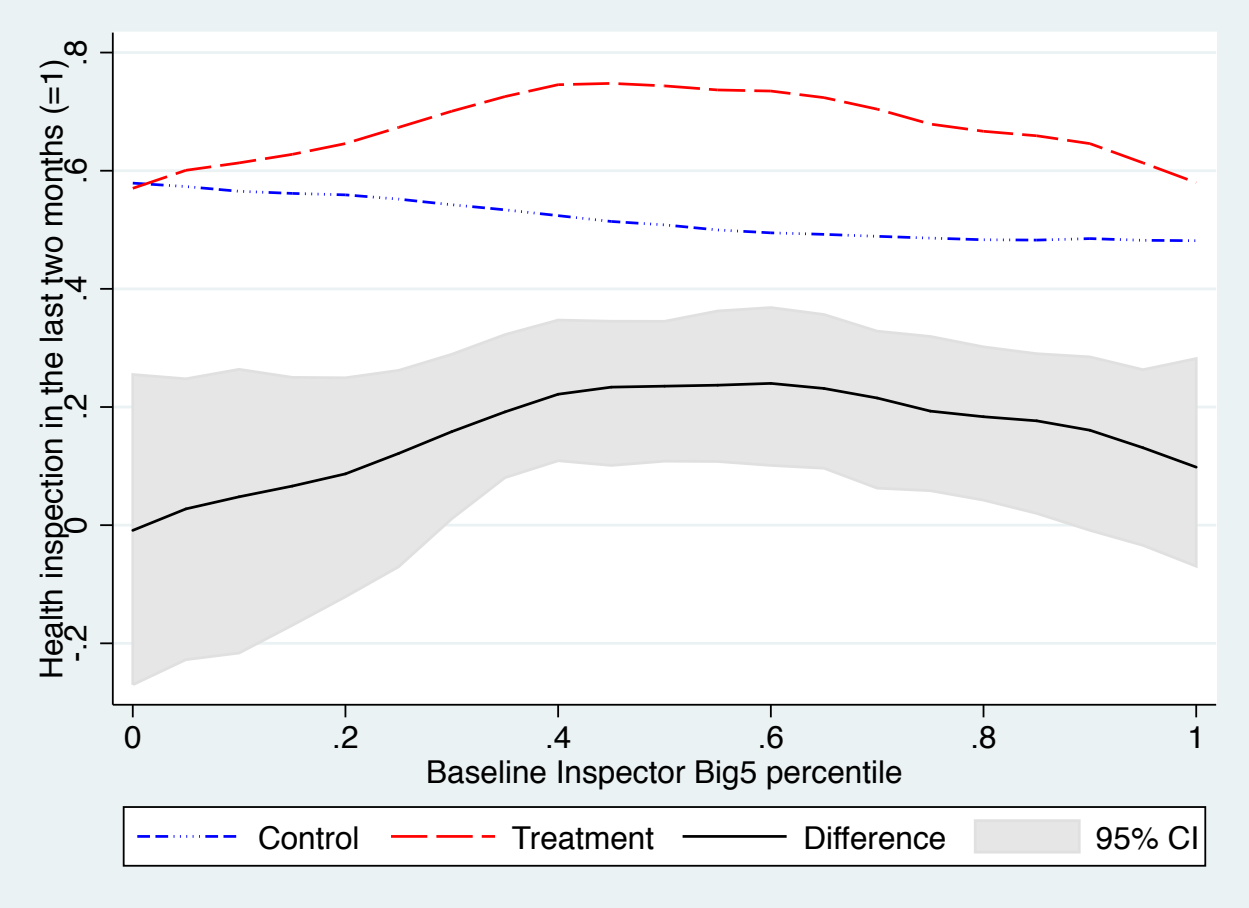


Figure 8 presents non-linear treatment effects of health inspector Big 5 index across the distribution of Big 5 personality. We can see that the effect in Table 6 is primarily being driven by those health inspectors in the middle of the Big 5 distribution. This fits the model presented in Section 4 in which we expect the effects of this intervention to be localized to those inspectors that were just below the marginal cutoff beforehand. Econometrically, this highlights not only that we must always be wary of generalizing our linear results. Practically, we find this point very intuitive and note that there are clear policy recommendations that accompany such effects, such as the gains to be had from targeting new interventions to

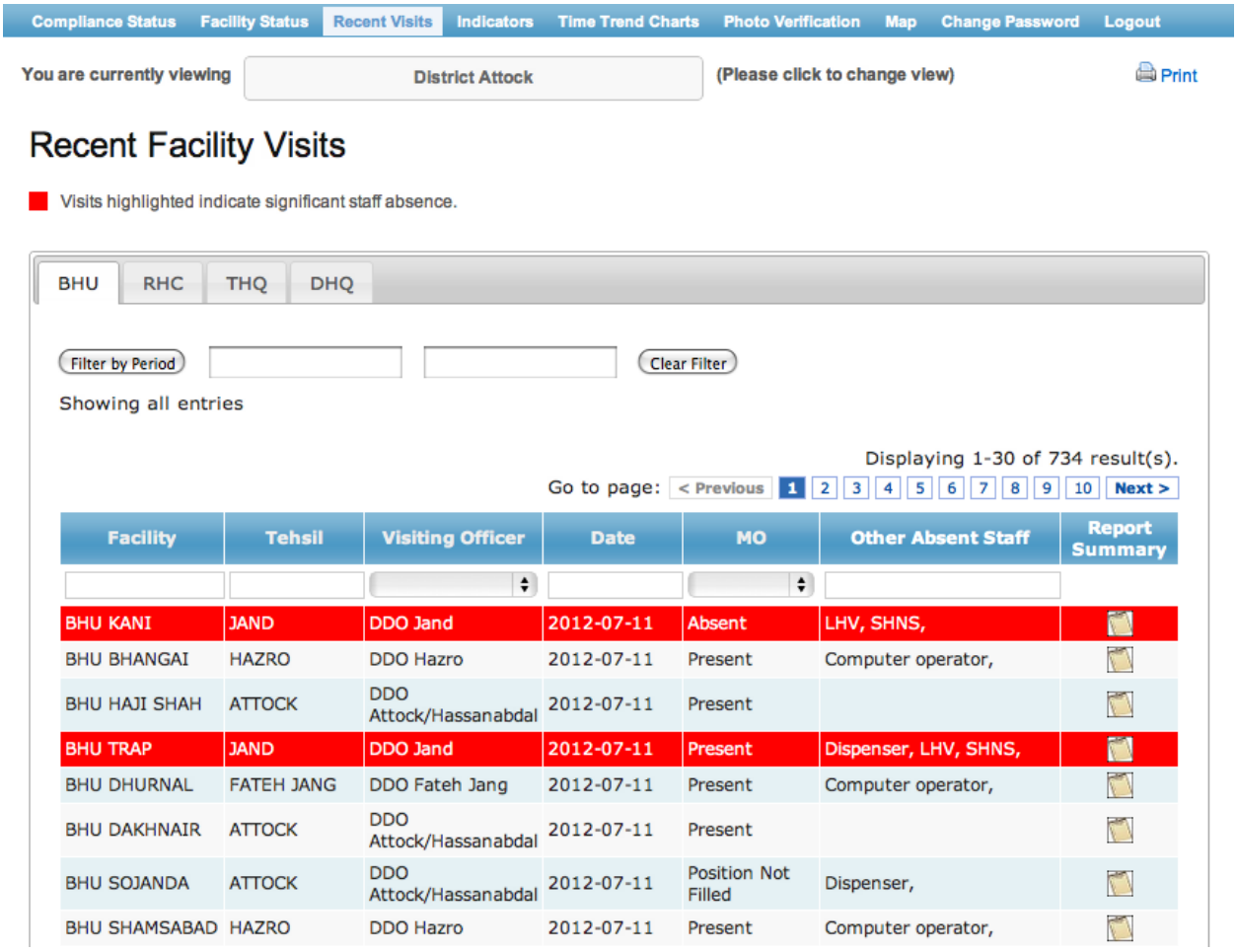
[18]Note that to test for robustness in our effects to the small number of district clusters in our analysis, we have conducted Fisher exact tests for all results. In all cases, the estimated p-value is as at least as significant as from OLS. Results available upon request.

those government workers in the middle of the personality distribution.

5.3 Heterogeneous Flagging Effects

Our experiment was designed to allow us a direct test of one mechanism that may create an increase in doctor attendance given increased monitoring—pressure from senior health officials following a negative report. Data from health inspections are aggregated and presented to senior health officials in each district of Punjab (EDOs, of which there are one per district) on an online dashboard. This dashboard is only visible to these senior health officials as well as to the Health Secretary for Punjab and the Director General of Health for Punjab. Figure 9 provides an example of a dashboard view visible to senior health officials.

Figure 9: Highlighting Underperforming Facilities to Test Mechanisms



Facility	Tehsil	Visiting Officer	Date	MO	Other Absent Staff	Report Summary
BHU KANI	JAND	DDO Jand	2012-07-11	Absent	LHV, SHNS,	
BHU BHANGAI	HAZRO	DDO Hazro	2012-07-11	Present	Computer operator,	
BHU HAJI SHAH	ATTOCK	DDO Attock/Hassanabdal	2012-07-11	Present		
BHU TRAP	JAND	DDO Jand	2012-07-11	Present	Dispenser, LHV, SHNS,	
BHU DHURNAL	FATEH JANG	DDO Fateh Jang	2012-07-11	Present	Computer operator,	
BHU DAKHNAIR	ATTOCK	DDO Attock/Hassanabdal	2012-07-11	Present		
BHU SOJANDA	ATTOCK	DDO Attock/Hassanabdal	2012-07-11	Position Not Filled	Dispenser,	
BHU SHAMSABAD	HAZRO	DDO Hazro	2012-07-11	Present	Computer operator,	

Table 7: Senior health official personality summary statistics (control districts)

	Mean	SD	P10	P50	P90	Obs
Big5 personality traits						
Big 5 index (z-score)	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 (z-score)	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
Prosocial behavior						
Prosocial (principal comp.)	0.22	1.56	-3.38	0.78	1.42	16
Did charity work in the past year (=1)	0.88	0.34	0.00	1.00	1.00	16
Has ever run for office (=1)	0.00	0.00	0.00	0.00	0.00	16
Has ever volunteered (=1)	0.88	0.34	0.00	1.00	1.00	16
Voted in the last NA election (=1)	0.75	0.45	0.00	1.00	1.00	16
Has ever donated blood (=1)	0.81	0.40	0.00	1.00	1.00	16
Goes to the mosque regularly (=1)	0.50	0.52	0.00	0.50	1.00	16
Believes people can be trusted (=1)	0.69	0.48	0.00	1.00	1.00	16

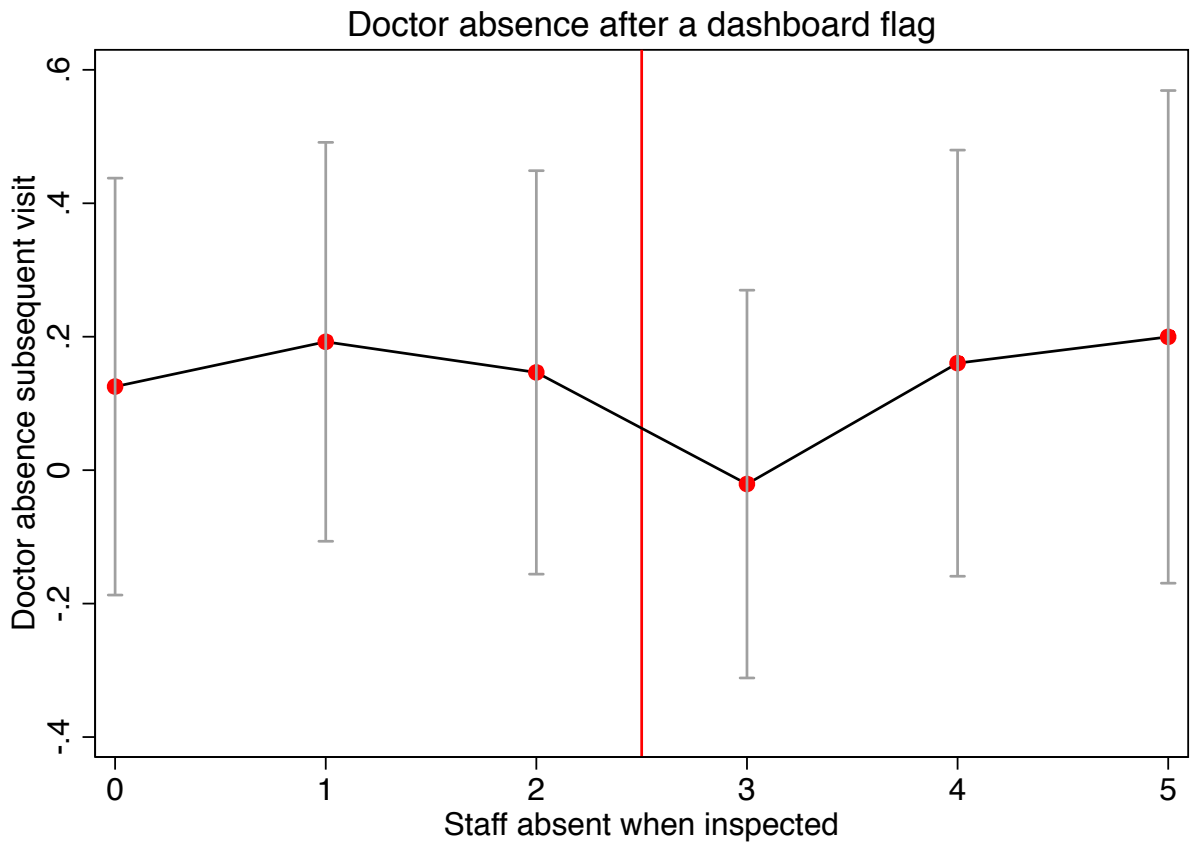
Notes: Sample: senior health officials in control districts that completed the personalities survey module, given during a single round after the final wave of clinic visits. All personality traits and public sector motivation variables measured on a one to five Likert scale unless otherwise indicated.

Table 7 presents the summary statistics of the senior health officials in Punjab. We see that they are very similar to summary statistics of both doctors and health inspectors. This is more evidence in the comparability of these measures across individuals.

To test whether actions by senior health officials affect subsequent absence, we directly manipulate the data on the dashboard to make certain facilities salient. Specifically, we highlight entries that find three or more staff to be absent in red on the dashboard. In Callen et al. (2013), the authors examine at length whether this manipulation affects subsequent doctor absence, finding consistent evidence that flagging facilities leads to decreased subsequent doctor absence. This is summarized in Figure 10, where we see a noticeable drop in subsequent doctor absence moving between 2 and 3 staff absent during a flagging window, exactly where the RD should have bite.

As was done in the pervious sub-section, we simply interact our personality measures with a dummy for highlighted entries ('Flagged'). Table 8 reports results from this test, limited to only the "discontinuity" sample of those facilities with 2 or 3 staff absent during the flagging window. Table A.5 verifies what we expect from the picture above for the index personality measures—that the drop is limited to right around the discontinuity, though it is sizable enough to come through in direction and magnitude on the entire sample. The results in Table 8 here are even stronger than previous experimental results. We see that a one standard deviation higher senior health official Big 5 index is associated with an over 40 percentage point decrease in the rate of doctor absence in a facility following the facility being flagged as underperforming on the dashboard. In this case, we define the window during which a flagging could occur prior to one of our unannounced visits as 15 to 45 days before our visit. This is because we know that senior health officials only looked at the web dashboard every week or two, so we wouldn't expect an immediate response, but if we go too far back, virtually every facility will become 'flagged' and we will lose variation. Note that these results are robust to the window selected as one would expect—too short of a window or too long of a window and the results disappear, as well as if the window moves

Figure 10: Flagging effects



Notes: Each point represents a coefficient from one regression of absence on a series of dummies for the maximum number of individuals absent at a facility in any visit during a flagging window. The regression includes district and survey wave fixed effects. 95 percent confidence intervals are shown, from standard errors clustered at the clinic level. Note clinics were flagged as underperforming if 3 or more of the 7 staff were absent in the last visit.

too close or to far away from our visit. The p-values of the significance of the coefficient on the Big 5 index for a wide range of windows are reported in Table A.6.

First and foremost, we take these results to be another strong validation of these personality measures in predicting performance, this time in the case of senior health officials. We thus see that personality measures can predict performance of not just low- and mid-level public employees, but also of very senior ones. And personality measure predicts performance with extremely large magnitudes in this case as well—flagging essentially eliminates doctor absence in subsequent visits for senior health officials one standard deviation above

Table 8: Differential clinic 'flagging' effects by senior health officer personality

	Doctor absent (=1)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PANEL A: Big 5 personality traits									
Clinic flagged as underperforming on dashboard			-0.146	-0.159	0.467	1.331	1.089	-1.012**	0.318
			(0.103)	(0.098)	(1.022)	(0.843)	(1.231)	(0.490)	(0.965)
Flagged x Big5 index				-0.402**					
				(0.200)					
Flagged x Agreeableness					-0.166				
					(0.278)				
Flagged x Conscientiousness						-0.359*			
						(0.202)			
Flagged x Extroversion							-0.322		
							(0.318)		
Flagged x Emotional stability								-0.361*	
								(0.205)	
Flagged x Openness									-0.157
									(0.326)
Mean of the dependent variable			0.480	0.480	0.480	0.480	0.480	0.480	0.480
# Observations			123	123	123	123	123	123	123
# Clinics			106	106	106	106	106	106	106
R-Squared			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		-0.146	-0.165	0.326	0.137	2.449	-0.418	-0.433	1.187
		(0.103)	(0.105)	(0.661)	(0.946)	(1.673)	(1.134)	(0.903)	(0.938)
Flagged x PSM index			-0.124						
			(0.169)						
Flagged x Attraction				-0.128					
				(0.180)					
Flagged x Civic duty					-0.065				
					(0.214)				
Flagged x Commitment						-0.700			
						(0.450)			
Flagged x Compassion							0.071		
							(0.292)		
Flagged x Self sacrifice								0.066	
								(0.205)	
Flagged x Social justice									-0.343
									(0.240)
Mean of the dependent variable		0.480	0.480	0.480	0.480	0.480	0.480	0.480	0.480
# Observations		123	123	123	123	123	123	123	123
# Clinics		106	106	106	106	106	106	106	106
R-Squared		0.204	0.208	0.207	0.204	0.217	0.204	0.204	0.219
PANEL B: Public service motivation									
Clinic flagged as underperforming on dashboard	-0.174	-0.155	-0.201	-0.146	-0.350	-0.302	-0.201	-0.349**	-0.097
	(0.122)	(0.122)	(0.211)	(0.103)	(0.282)	(0.211)	(0.211)	(0.162)	(0.175)
Flagged x Prosocial (principal comp.)		0.022							
		(0.056)							
Flagged x Did charity work in the past year (=1)			0.080						
			(0.240)						
Flagged x Has even run for office (=1)				(dropped)					
				.					
Flagged x Has ever volunteered (=1)					0.241				
					(0.302)				
Flagged x Voted in the last NA election (=1)						0.206			
						(0.241)			
Flagged x Has ever donated blood (=1)							0.080		
							(0.240)		
Flagged x Goes to the mosque regularly (=1)								0.334	
								(0.203)	
Flagged x Believes people can be trusted (=1)									-0.178
									(0.235)
Mean of the dependent variable	0.530	0.530	0.480	0.480	0.480	0.480	0.480	0.480	0.530
# Observations	100	100	123	123	123	123	123	123	100
# Clinics	87	87	106	106	106	106	106	106	87
R-Squared	0.188	0.190	0.205	0.204	0.209	0.209	0.205	0.223	0.194

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. Clinics were flagged as underperforming if 3 or more of the 7 staff were absent in the last visit. All columns restrict the sample to those clinics where only 2 or 3 staff were absent (up to 7 staff can be marked absent).

the mean in the Big 5 index! We also take these results as a hint at the type of mechanisms that may be at play here. Though distinct from previous results that rely on doctor and health inspector personality, these results suggest that personality may lead to better performance in the face of a policy change aimed at increasing information through the most straightforward mechanism—response to salient information. Of course, we can't speak to the exact way that these senior health officials acted on salient information to decrease absence in subsequent months. Anecdotally, we know that it is unlikely that they took serious action against delinquent workers, but it is more likely that they simply gave such workers a stern phone call or invited them to a meeting in the district headquarters.

6 Conclusion

We provide three major sets of results that support the relevance of personalities for policy adoption. First, Big 5 characteristics and Public Sector Motivation positively predict doctor attendance and negatively predict whether doctors collude with inspectors to falsify reports. Second, smart phone monitoring has the largest impact on health inspectors with high Big 5 personality measures. Last, senior health officials with high Big 5 aggregate z-scores are most likely to respond to a report of underperforming clinic as measured by improved subsequent performance at the facility.

We point to four implications of these results. First, the size of our correlations suggest that psychometric measures such as these could potentially provide useful diagnostics in public sector hiring, training, and promotion decisions. Second, all of the public employees in this studies faced broadly the same incentives, yet performed very differently. Our results suggest that part of this difference may relate to differences in intrinsic motivation. Third, research shows that personality traits are malleable. This may expand the set of potential policy interventions to improve public service delivery. Last, research documents that the profile of applicants to public sector jobs is influenced by adjustable features of the position,

including the wage. Our results suggest that changing the personality profile of public servants can improve service delivery, without adjusting other incentives.

References

Almlund, Mathilde, Angela Lee Duckworth, James J. Heckman, and Tim D. Katz, *Personality Psychology and Economics*, Vol. 4, Elsevier, NBER Working Paper 16822.

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.

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

Bo, Ernesto Dal, Frederico Finan, and Martin 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.*

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.

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

Heckman, James J., "Integrating Personality Psychology into Economics," Technical Report 17378, NBER 2011.

_ , 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.

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.

John, Oliver P., Laura P. Naumann, and Christopher J. Soto, *Handbook of Personality*, The Guilford Press, 2008.

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.

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.

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

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. Unpublished manuscript.

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.

A Appendix

A.1 Appendix tables

Table A.1: Doctor personality and doctor attendance

	Doctor Present (=1)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PANEL A: Personality traits								
Big 5 index (z-score)			0.037					
			(0.034)					
Agreeableness				0.009				
				(0.036)				
Conscientiousness					0.098**			
					(0.047)			
Extroversion						0.093*		
						(0.052)		
Emotional stability							0.037	
							(0.036)	
Openness								-0.043
								(0.059)
Mean of dependent variable			0.493	0.493	0.493	0.493	0.493	0.493
# Clinics			190	190	190	190	190	190
# Observations			479	479	479	479	479	479
R-Squared			0.192	0.190	0.197	0.195	0.191	0.190
PANEL B: Public service motivation								
PSM index (z-score)		0.074**						
		(0.036)						
Attraction			0.048					
			(0.042)					
Civic duty				0.115**				
				(0.051)				
Commitment					0.060			
					(0.052)			
Compassion						0.015		
						(0.053)		
Self Sacrifice							0.089**	
							(0.042)	
Social justice								0.047
								(0.038)
Mean of dependent variable		0.493	0.493	0.493	0.493	0.493	0.493	0.493
# Clinics		190	190	190	190	190	190	190
# Observations		479	479	479	479	479	479	479
R-Squared		0.196	0.192	0.199	0.192	0.190	0.197	0.192
PANEL C: Prosocial behavior								
Prosocial (principal comp.)	0.007							
	(0.021)							
Did charity work in the past year (=1)		-0.010						
		(0.080)						
Has ever run for office (=1)			0.176					
			(0.128)					
Has ever volunteered (=1)				-0.028				
				(0.074)				
Voted in the last NA election (=1)					-0.026			
					(0.058)			
Has ever donated blood (=1)						0.081		
						(0.052)		
Goes to the mosque regularly (=1)							0.045	
							(0.052)	
Believes people can be trusted (=1)								0.041
								(0.053)
Mean of dependent variable	0.496	0.493	0.493	0.493	0.493	0.493	0.496	0.493
# Clinics	182	190	190	189	190	190	182	190
# Observations	456	479	479	477	479	479	456	479
R-Squared	0.198	0.189	0.191	0.191	0.190	0.193	0.199	0.191

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the clinic level reported in parentheses. All regressions include Tehsil (county) and survey wave fixed effects. Sample: control district clinics for which doctor personality data is available and a doctor is posted.

Table A.2: Doctor personality and estimated doctor-inspector collusion

	Doctor personality and estimated doctor-inspector collusion (=1)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PANEL A: Personality traits								
Big 5 index (z-score)			-0.104***					
			(0.031)					
Agreeableness				-0.125***				
				(0.043)				
Conscientiousness					-0.112***			
					(0.039)			
Extroversion						-0.131***		
						(0.047)		
Emotional stability							-0.092***	
							(0.032)	
Openness								-0.064
								(0.068)
Mean of dependent variable			0.094	0.094	0.094	0.094	0.094	0.094
# Clinics			245	245	245	245	245	245
# Observations			245	245	245	245	245	245
R-Squared			0.391	0.392	0.374	0.376	0.370	0.344
PANEL B: Public service motivation								
PSM index (z-score)		-0.149***						
		(0.036)						
Attraction			-0.107***					
			(0.037)					
Civic duty				-0.114***				
				(0.041)				
Commitment					-0.152***			
					(0.048)			
Compassion						-0.141***		
						(0.046)		
Self Sacrifice							-0.136***	
							(0.039)	
Social justice								-0.100**
								(0.038)
Mean of dependent variable		0.094	0.094	0.094	0.094	0.094	0.094	0.094
# Clinics		245	245	245	245	245	245	245
# Observations		245	245	245	245	245	245	245
R-Squared		0.433	0.379	0.382	0.393	0.383	0.393	0.369
PANEL C: Prosocial behavior								
Prosocial (principal comp.)	-0.041**							
	(0.019)							
Did charity work in the past year (=1)		-0.111						
		(0.074)						
Has ever run for office (=1)			-0.150					
			(0.111)					
Has ever volunteered (=1)				-0.077				
				(0.064)				
Voted in the last NA election (=1)					0.019			
					(0.042)			
Has ever donated blood (=1)						-0.104**		
						(0.051)		
Goes to the mosque regularly (=1)							-0.014	
							(0.046)	
Believes people can be trusted (=1)								0.012
								(0.040)
Mean of dependent variable	0.091	0.091	0.095	0.094	0.094	0.094	0.093	0.095
# Clinics	232	241	243	244	244	244	236	243
# Observations	232	241	243	244	244	244	236	243
R-Squared	0.360	0.350	0.342	0.346	0.340	0.357	0.338	0.350

Notes: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Standard errors clustered at the clinic level reported in parentheses. All regressions include Tehsil (county) and survey wave fixed effects. Sample: treatment district clinics for which doctor personality data is available and 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 73 visits).

Table A.3: Health inspector personality and inspections

	Inspection in the last two months (=1)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PANEL A: Personality traits								
Big 5 index (z-score)			-0.063					
			(0.049)					
Agreeableness				-0.047				
				(0.061)				
Conscientiousness					-0.100*			
					(0.059)			
Extroversion						-0.093		
						(0.073)		
Emotional stability							-0.102	
							(0.061)	
Openness								0.038
								(0.078)
Mean of dependent variable			0.511	0.511	0.511	0.511	0.511	0.511
# Inspectors			46	46	46	46	46	46
# Observations			523	523	523	523	523	523
R-Squared			0.181	0.179	0.182	0.182	0.183	0.178
Panel B: Public service motivation								
PSM index (z-score)		-0.021						
		(0.058)						
Attraction			-0.027					
			(0.065)					
Civic duty				0.017				
				(0.060)				
Commitment					-0.016			
					(0.087)			
Compassion						-0.095		
						(0.114)		
Self Sacrifice							-0.002	
							(0.044)	
Social justice								-0.031
								(0.080)
Mean of dependent variable		0.495	0.495	0.495	0.495	0.495	0.495	0.495
# Inspectors		47	47	47	47	47	47	47
# Observations		539	539	539	539	539	539	539
R-Squared		0.199	0.200	0.199	0.199	0.202	0.199	0.199
PANEL C: Prosocial behavior								
Prosocial (principal comp.)	0.056**							
	(0.026)							
Did charity work in the past year (=1)		0.245						
		(0.165)						
Has ever run for office (=1)			-0.053					
			(0.050)					
Has ever volunteered (=1)				0.189*				
				(0.106)				
Voted in the last NA election (=1)					-0.008			
					(0.062)			
Has ever donated blood (=1)						0.085		
						(0.055)		
Goes to the mosque regularly (=1)							0.129**	
							(0.062)	
Believes people can be trusted (=1)								-0.052
								(0.061)
Mean of dependent variable	0.495	0.495	0.495	0.495	0.495	0.495	0.495	0.495
# Inspectors	47	47	47	47	47	47	47	47
# Observations	539	539	539	539	539	539	539	539
R-Squared	0.211	0.208	0.199	0.209	0.199	0.202	0.208	0.200

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard errors clustered at the Tehsil level reported in parentheses. All regressions include district and survey wave fixed effects. Sample: control district Basic Health Units (BHUs) for which health inspector (DDO) personality data is available.

Table A.4: Health inspector personality and estimated doctor-inspector collusion

	Doctor-DDO collusion (=1)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PANEL A: Personality traits								
Big 5 index (z-score)			0.022					
			(0.039)					
Agreeableness				-0.011				
				(0.056)				
Conscientiousness					0.017			
					(0.045)			
Extroversion						0.079*		
						(0.047)		
Emotional stability							-0.003	
							(0.027)	
Openness								-0.011
								(0.040)
Mean of dependent variable			0.091	0.091	0.091	0.091	0.091	0.091
# Inspectors			46	46	46	46	46	46
# Observations			252	252	252	252	252	252
R-Squared			0.178	0.177	0.177	0.184	0.177	0.177
PANEL B: Public service motivation								
PSM index (z-score)		-0.074***						
		(0.028)						
Attraction			-0.115*					
			(0.058)					
Civic duty				0.016				
				(0.038)				
Commitment					-0.124***			
					(0.038)			
Compassion						-0.057		
						(0.049)		
Self Sacrifice							-0.068**	
							(0.030)	
Social justice								-0.053
								(0.038)
Mean of dependent variable		0.094	0.094	0.094	0.094	0.094	0.094	0.094
# Inspectors		47	47	47	47	47	47	47
# Observations		254	254	254	254	254	254	254
R-Squared		0.196	0.193	0.183	0.204	0.187	0.189	0.187
PANEL C: Prosocial behavior								
Prosocial (principal comp.)	-0.041							
	(0.028)							
Did charity work in the past year (=1)		-0.283**						
		(0.123)						
Has ever run for office (=1)			0.000					
			(0.000)					
Has ever volunteered (=1)				0.041				
				(0.038)				
Voted in the last NA election (=1)					-0.089			
					(0.065)			
Has ever donated blood (=1)						-0.091**		
						(0.040)		
Goes to the mosque regularly (=1)							-0.057	
							(0.047)	
Believes people can be trusted (=1)								-0.004
								(0.041)
Mean of dependent variable	0.091	0.091	0.093	0.091	0.091	0.091	0.091	0.093
# Inspectors	46	46	45	46	46	46	46	45
# Observations	252	252	236	252	252	252	252	236
R-Squared	0.187	0.200	0.185	0.178	0.182	0.186	0.181	0.185

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard errors clustered at the tehsil (health inspector) level reported in parentheses. All regressions include Tehsil (county) and survey wave fixed effects. Sample: treatment district clinics for which health inspector personality data is available. 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 73 visits).

Table A.5: Differential clinic flagging effects by senior health official personality

	Doctor absent (=1)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Clinic flagged as underperforming on dashboard	-0.100	-0.146	-0.094	-0.159	-0.098	-0.165	-0.130*	-0.155
	(0.067)	(0.103)	(0.067)	(0.098)	(0.070)	(0.105)	(0.077)	(0.122)
Flagged x Senior health official Big5 index			-0.118	-0.402**				
			(0.131)	(0.200)				
Flagged x Senior health official PSM index					0.016	-0.124		
					(0.108)	(0.169)		
Flagged x Senior health official prosocial princ. comp.							0.002	0.022
							(0.042)	(0.056)
Mean of the dependent variable	0.521	0.480	0.521	0.480	0.521	0.480	0.559	0.530
# Observations	326	123	326	123	326	123	281	100
# Clinics	228	106	228	106	228	106	201	87
R-Squared	0.114	0.204	0.117	0.231	0.114	0.208	0.093	0.190
Sample	Full	Discontinuity	Full	Discontinuity	Full	Discontinuity	Full	Discontinuity

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. Clinics were flagged as underperforming if 3 or more of the 7 staff were absent in the last visit. Columns 2 and 4 restrict the sample to those clinics where only 2 or 3 staff were absent (up to 7 staff can be marked absent). We call this sample the "discontinuity" sample.

Table A.6: Robustness to different windows for flagging

	Length of window							
Window begins at	5	10	15	20	25	30	35	40
t-0	0.12	0.94	0.73	0.25	0.06	0.09	0.11	0.45
t-5	0.16	0.12	0.67	0.20	0.16	0.18	0.62	0.45
t-10	0.29	0.31	0.05	0.05	0.07	0.29	0.21	0.43
t-15	0.00	0.00	0.00	0.00	0.03	0.04	0.15	0.21
t-20	0.00	0.01	0.05	0.18	0.26	0.44	0.53	0.55
t-25	0.51	0.73	0.93	0.89	0.58	0.64	0.78	0.58

Notes: Each cell reports the p-value of a test of the null hypothesis that the interaction term of senior health official Big 5 personality index and 'flagging' from the regression in Table 8 is equal to zero, where the indicated flagging window was used. The cell for window begins at: t-15 and length of window: 30 coincides with the results in Table 8.

A.2 Personalities Survey Instrument—Urdu

نام _____ عہدہ: _____ یونین کونسل نمبر: _____

BHU کا نام: _____ HMIS سہولت کا کوڈ: _____

Part 3

Medical Officer

(Self Reporting Section)

LUMS کی موجودہ اسٹڈی کے اس حصے میں ہم یہ کوشش کر رہے ہیں کہ BHU میں تعینات صحت کے افسران سے کام کے حوالے سے ان کی تسلی اور پوسٹ پر قائم رہنے پر اثر انداز ہونے والے محرکات کے بارے میں باقاعدہ طریقہ سے معلومات جمع کریں۔ ہم آپ کے بہت شکر گزار ہوں گے کہ آپ کچھ وقت نکال کر اس لفافے میں موجود شیٹ کو پُر کر دیں اور لفافے میں ڈال کر انٹرویور کو واپس دے دیں۔ ہم آپ کی یاددہانی کے لیے بتاتے چلیں کہ باقی سروے کی طرح اس حصے میں بھی آپ کے تمام جوابات کو ہماری تحقیقاتی ٹیم صیغہ راز میں رکھے گی اور اس کو محکمہ صحت کے کسی فرد کے ساتھ شیئر نہیں کیا جائے گا۔ علاوہ ازیں پہلے کی طرح آپ کی شمولیت رضاکارانہ ہوگی۔

سوالنامہ پر کرنے کے لیے ہدایات:

1- ہر بیان کو احتیاط سے پڑھیے اور جس جواب سے آپ متفق ہیں، اس جواب کے گرد دائرہ لگائیں۔

اگر آپ بیان سے "بالکل متفق نہیں" تو (1) کے گرد دائرہ لگائیے۔

اگر آپ بیان سے "زیادہ تر متفق نہیں" تو (2) کے گرد دائرہ لگائیے۔

اگر آپ بیان کے بارے میں "غیر جانبدار ہیں" تو (3) کے گرد دائرہ لگائیے۔

اگر آپ بیان سے "زیادہ تر متفق ہیں" تو (4) کے گرد دائرہ لگائیے۔

اگر آپ بیان سے "مکمل متفق ہیں" تو (5) کے گرد دائرہ لگائیے۔

2- اس ٹیسٹ میں صحیح اور غلط کا کوئی تصور نہیں ہے اور نہ ہی اس کو حل کرنے کے لیے آپ کے ماہر ہونے کی ضرورت ہے۔ ایمانداری سے ممکنہ حد تک ٹھیک جواب دیں۔ اپنی رائے کو جہاں تک ممکن ہو سکے احتیاط اور سچائی سے لکھیں۔ ہر سوال کا جواب ضرور دیجیے اور یہ یقین کر لیں کہ آپ نے ہر جواب کے لیے صحیح جگہ پر دائرہ لگایا ہے۔ ٹیسٹ کے دوران اگر آپ سے غلطی ہو جائے یا آپ کا ذہن بدل جائے، تو اپنے جواب کو مت مٹائیں۔ اس کے لیے غلط جواب پر (x) کا نشان لگائیں اور ٹھیک جواب کے گرد دائرہ لگائیے۔

Section-1

نمبر شمار	بیانات	مکمل غیر متفق	زیادہ تر غیر متفق	غیر جانبدار	زیادہ تر متفق	مکمل متفق
1	سیاست ایک برا لفظ ہے	1	2	3	4	5
2	میں ایسے منتخب عوامی نمائندے کا احترام کرتا ہوں جو اچھی سوچ کو قانون میں تبدیل کر سکیں	1	2	3	4	5
3	منتخب عوامی نمائندے کا اخلاقی رویہ اتنا ہی اہم ہے جتنی ان کی قابلیت۔	1	2	3	4	5
4	کچھ حاصل کرنے کے لیے کچھ کھونا کی بنیاد پر سیاسی لین دین مجھے متاثر نہیں کرتا	1	2	3	4	5
5	میں سیاست دانوں کی کچھ خاص پرواہ نہیں کرتا۔	1	2	3	4	5
6	لوگ عوام کے فائدے کی بات تو کرتے ہیں مگر حقیقت میں وہ صرف ذاتی فائدے کے بارے میں سوچ رہے ہوتے ہیں۔	1	2	3	4	5
7	میرے لیے اپنے علاقہ میں ہونے والے واقعات میں شدت سے دلچسپی لینا دشوار ہوتا ہے۔	1	2	3	4	5
8	میں اپنے علاقے کے لیے بے لوث ہو کر کام کرتا ہوں۔	1	2	3	4	5
9	بامعنی عوامی خدمت میرے لیے بہت اہمیت رکھتی ہے۔	1	2	3	4	5
10	میں اس بات کو فوقیت دوں گا کہ منتخب عوامی نمائندے علاقے کے فائدے کے لیے کام کریں، چاہے وہ میرے ذاتی مفاد میں نہ ہوں۔	1	2	3	4	5
11	ایک سرکاری ملازم کی عوامی ذمہ داری اپنے افسران کے لیے وفاداری سے زیادہ اہم ہے۔	1	2	3	4	5
12	میں عوام کی خدمت کو اپنی سماجی ذمہ داری سمجھتا ہوں۔	1	2	3	4	5
13	مجھے یقین ہے کہ بہت سے عوامی مقاصد ایسے ہیں جن کو سر کرنے کی ضرورت ہے۔	1	2	3	4	5
14	میں اس بات پر یقین نہیں رکھتا کہ حکومت معاشرے کو مزید منصفانہ بنانے کے لیے کچھ کر سکتی ہے۔	1	2	3	4	5
15	اگر کوئی بھی گروہ معاشرے کی خوشحالی سے محروم رہ جاتا ہے تو پھر ہم بُرے حالات میں ہی رہیں گے۔	1	2	3	4	5
16	میں اس دنیا کو زیادہ منصفانہ جگہ بنا دینے کے لیے اپنی توانائی کا ہر ذرہ لگا دینے کے لئے تیار ہوں۔	1	2	3	4	5
17	میں دوسروں کے حقوق کے لیے آواز اٹھانے سے نہیں ڈرتا چاہے میرا مذاق ہی کیوں نہ اڑایا جائے۔	1	2	3	4	5
18	جب سرکاری ملازمین حلف اٹھاتے ہیں تو میں سمجھتا ہوں کہ وہ جو ذمہ داریاں اٹھانے کے لیے تیار ہوتے ہیں جو دوسرے شہریوں سے متوقع نہیں کی جا سکتی ہیں۔	1	2	3	4	5
19	میں اپنی ملکی ذمہ داریاں پوری کرنے کے لیے کسی بھی حد تک جا سکتا ہوں۔	1	2	3	4	5
20	سرکاری ملازمت شہریت کا سب سے اونچا درجہ ہے۔	1	2	3	4	5

نمبر شمار	بیانات	مکمل غیر متفق	زیادہ تر غیر متفق	غیر جانبدار	زیادہ تر متفق	مکمل متفق
21	مجھے یقین ہے کہ تمام لوگ چاہے کتنے ہی مصروف کیوں نہ ہوں۔ اُن کی اخلاقی ذمہ داری ہے کہ وہ سماجی معاملات کو اچھے طریقے سے نبھائیں۔	1	2	3	4	5
22	میری ذمہ داری ہے کہ میں غریبوں کی دیکھ بھال کروں۔	1	2	3	4	5
23	"کام"، "عزت" اور "ملک"، کے الفاظ میرے دل کی گہرائیوں میں چھپے احساسات جگاتے ہیں۔	1	2	3	4	5
24	میری ذمہ داری ہے کہ لوگوں کے باہمی انحصار سے پیدا ہونے والے مسائل کو حل کروں۔	1	2	3	4	5
25	میں محروم لوگوں کی اذیت سے کبھی کبھار ہی متاثر ہوتا ہوں۔	1	2	3	4	5
26	بہت سے سوشل پروگرام بہت اہم ہیں، جن کے بغیر گزارہ نہیں ہوسکتا۔	1	2	3	4	5
27	جب میں لوگوں کو پریشانی میں دیکھتا ہوں تو میرے لیے اپنے جذبات کو قابو میں رکھنا مشکل ہوجاتا ہے۔	1	2	3	4	5
28	میرے نزدیک دوسروں کی فلاح کے لیے کام کرنا حبُ الوطنی کا اظہار ہے۔	1	2	3	4	5
29	میں اُن لوگوں کی فلاح کے لیے شاذ و نادر ہی سوچتا ہوں جن کو ذاتی طور پر نہیں جانتا۔	1	2	3	4	5
30	روزمرہ کے واقعات مجھے بارہا یہ احساس دلاتے ہیں کہ ہم ایک دوسرے پر کس حد تک انحصار کرتے ہیں۔	1	2	3	4	5
31	مجھے اُن لوگوں سے کوئی خاص ہمدردی نہیں ہوتی جو اپنی ضروریات کو پورا کرنے کے لیے پہلا قدم اُٹھانے سے بھی گریز کرتے ہیں۔	1	2	3	4	5
32	میں بہت کم عوامی پروگراموں کی مکمل طور پر حمایت کرتا ہوں	1	2	3	4	5
33	معاشرے میں تبدیلی لانا میرے نزدیک زیادہ اہمیت رکھتا ہے، نسبت ذاتی کامیابیوں کے۔	1	2	3	4	5
34	میں فرائض کو ذاتی کاموں پر فوقیت دیتا ہوں۔	1	2	3	4	5
35	میرے نزدیک مالی طور پر مضبوط ہونا زیادہ اہمیت رکھتا ہے، نسبت اچھے کام کرنے کے۔	1	2	3	4	5
36	میں زیادہ تر جن مقاصد کے لیے کام کرتا ہوں وہ میرے ذاتی مفاد سے بڑھ کر اہمیت رکھتے ہیں۔	1	2	3	4	5
37	شہریوں کی خدمت کرنا میرے لیے باعث اطمینان ہوتا ہے چاہے اس کا کوئی معاوضہ مجھے نہ دے۔	1	2	3	4	5
38	میرے خیال میں لوگوں کو معاشرے کو اُس سے زیادہ دینا چاہیے جتنا وہ اس سے لیتے ہیں۔	1	2	3	4	5
39	میں اُن چند لوگوں میں سے ہوں جو دوسروں کی مدد کرنے کے لیے ذاتی نقصان اُٹھانے کیلئے بھی تیار ہوتے ہیں۔	1	2	3	4	5
40	میں معاشرے کی بھلائی کے لیے بڑی سے بڑی قربانی دینے کو بھی تیار ہوں۔	1	2	3	4	5

Section-2

نمبر شمار	بیانات	مکمل غیر متفق	زیادہ تر غیر متفق	غیر جانبدار	زیادہ تر متفق	مکمل متفق
1	میں ہر کام کی منصوبہ بندی پہلے ہی کرتا ہوں۔	1	2	3	4	5
2	میں فوراً فیصلہ کر لیتا ہوں۔	1	2	3	4	5
3	میں معمول میں بچت کرتا ہوں۔	1	2	3	4	5
4	جب میں اپنے کام سے دور ہوتا ہوں تو میں اپنے کام کی طرف واپس جانے کے لیے بے چین ہوتا ہوں۔	1	2	3	4	5
5	میں ایسے بہت سے مواقع کے بارے میں سوچ سکتا ہوں جب میں اپنا کام دلجمعی سے کر رہا تھا مگر دوسرے کام کرنا چھوڑ دیتے تھے	1	2	3	4	5
6	میں مشکل پروجیکٹ میں بھی اپنا کام جاری رکھتا، اس صورت میں بھی جب دوسرے میری مخالفت کرتے۔	1	2	3	4	5
7	میں ایک ہی وقت میں بہت سارے کام کرنا پسند کرتا ہوں۔	1	2	3	4	5
8	میں بہت سے پروجیکٹس کے حصوں کو مکمل کرنے کی نسبت ہر روز ایک پروجیکٹ کو مکمل کرنا پسند کرتا ہوں۔	1	2	3	4	5
9	میں یقین رکھتا ہوں کہ کسی نئے کام کو شروع کرنے سے پہلے پرانے کام کو مکمل کرنا بہتر ہے۔	1	2	3	4	5
10	یہ جاننا مشکل ہے کہ میرے اصل دوست کون ہیں۔	1	2	3	4	5
11	میں اُس کام کو کرنے کی کوشش نہیں کرتا جس کے بارے میں مجھے یقین نہ ہو۔	1	2	3	4	5
12	عمومی طور پر کہا جا سکتا ہے کہ اس علاقے کے لوگ ایماندار ہیں اور ان پر اعتبار کیا جا سکتا ہے۔	1	2	3	4	5
13	ایک آدمی خطرات مول لے کر امیر ہو سکتا ہے۔	1	2	3	4	5
14	اگر آنے والے ہفتے میں آپ کو وراثت میں یا پھر ایسے ہی ایک بڑی رقم دی جائے تو کیا پھر بھی صحت کے محکمہ کے ساتھ کام جاری رکھیں گے؟	1	2	3	4	5
15	آپ کو کتنی رقم دینے پر آپ نوکری چھوڑنے یا ریٹائر ہونے پر تیار ہو جائیں گے؟	_____ روپے				
16	اگر کسی شخص کو آپ کا بٹوہ ملے جس میں 2 ہزار روپے ہوں تو اس بات کا کتنا امکان ہے کہ وہ بٹوہ آپ کو تمام رقم سمیت واپس ہوگا؟					
	اگر یہ بٹوہ ملتا ہے:	بہت زیادہ امکان	کچھ امکان	غیر جانبدار	امکان نہیں	بالکل امکان نہیں
	16a- ہمسایہ کو	1	2	3	4	5
	16b- پولیس والے کو	1	2	3	4	5
	16c- کسی اجنبی کو	1	2	3	4	5

Section-3

نمبر شمار	بیانات	مکمل غیر متفق	زیادہ تر غیر متفق	غیر جانبدار	زیادہ تر متفق	مکمل متفق
1	میں پریشان حال نہیں ہوں۔	1	2	3	4	5
2	میں بہت سے لوگوں کے درمیان رہنا پسند کرتا/پسند کرتی ہوں۔	1	2	3	4	5
3	میں جاگتی آنکھوں خواب دیکھنے میں اپنا وقت ضائع کرنا پسند نہیں کرتا/ کرتی۔	1	2	3	4	5
4	میں اپنے ہر ملنے والے سے خوش اخلاقی سے پیش آنے کی کوشش کرتا/ کرتی ہوں۔	1	2	3	4	5
5	میں اپنی چیزیں صاف ستھری رکھتا/رکھتی ہوں۔	1	2	3	4	5
6	اکثر اوقات میں اپنے آپ کو دوسروں سے کمتر سمجھتا/ سمجھتی ہوں۔	1	2	3	4	5
7	میں آسانی سے ہنس لیتا ہوں/لیتی ہوں۔	1	2	3	4	5
8	جب کوئی کام کرنے کا صحیح طریقہ مجھے معلوم ہو جاتا ہے،تو میں اس پر جم جاتا ہوں۔	1	2	3	4	5
9	میری اکثر اپنے خاندان اور اپنے ساتھ کام کرنے والوں سے ٹوٹو میں میں ہو جاتی ہے۔	1	2	3	4	5
10	میں اپنے کام کی رفتاریوں متعین کرتا ہوں/ کرتی ہوں کہ سب کام وقت پر کر سکوں۔	1	2	3	4	5
11	بعض اوقات شدید ذہنی دباؤ ہو،تو مجھے محسوس ہوتا ہے کہ میرا وجود ٹوٹ کر بکھر جائے گا۔	1	2	3	4	5
12	میں اپنے آپ کو خوش دل طبیعت کا مالک نہیں سمجھتا/ سمجھتی۔	1	2	3	4	5
13	آرٹ اور قدرت کے نمونے مجھے محصور کر دیتے ہیں۔	1	2	3	4	5
14	بعض لوگوں کا خیال ہے کہ میں خود غرض اور انا پرست ہوں۔	1	2	3	4	5
15	میں زیادہ منظم شخص نہیں ہوں۔	1	2	3	4	5
16	میں شاذ و نادر ہی تنہائی یا افسردگی محسوس کرتا/ کرتی ہوں۔	1	2	3	4	5
17	مجھے لوگوں سے بات چیت کر کے واقعی لطف آتا ہے۔	1	2	3	4	5
18	میرے خیال میں طالب علموں کا ممتاز مقررین کو سننا، انہیں اُلجھا اور بھٹکا سکتا ہے۔	1	2	3	4	5
19	میں دوسروں سے مقابلہ کرنے کی بجائے اُن سے تعاون کرنے کو ترجیح دوں گا/دوں گی۔	1	2	3	4	5
20	میں وہ تمام کام جو میرے سپرد کئے گئے ہوں، اپنے ضمیر کے مطابق کرنے کی کوشش کرتا/ کرتی ہوں۔	1	2	3	4	5
21	میں اکثر اوقات ذہنی تناؤ اور گھبراہٹ محسوس کرتا/ کرتی ہوں۔	1	2	3	4	5
22	مجھے اکثر ولولہ انگیز صورت حال کی تمنا ہوتی ہے۔	1	2	3	4	5
23	شاعری کا مجھ پر بہت ہی کم یا نا ہونے کے برابر اثر ہوتا ہے۔	1	2	3	4	5
24	میں دوسرں کی نیت کے بارے میں بدگمان اور شکی ہوں۔	1	2	3	4	5
25	میرے مقاصد بہت واضح ہیں اور میں ان کے لئے بہت منظم طریقے سے کام کرتا ہوں/ کرتی ہوں۔	1	2	3	4	5

نمبر شمار	بیانات	مکمل غیر متفق	زیادہ تر غیر متفق	غیر جانبدار	زیادہ تر متفق	مکمل متفق
26	بعض اوقات میں خود کو ایک کوڑی کا بھی نہیں سمجھتا/ سمجھتی۔	1	2	3	4	5
27	عموماً میں اکیلے ہی کام کرنے کو ترجیح دیتا ہوں/دیتی ہوں۔	1	2	3	4	5
28	میں اکثر نئے اور غیر ملکی کھانے آزماتا ہوں/ آزماتی ہوں۔	1	2	3	4	5
29	مجھے یقین ہے اگر آپ موقع دیں تو دوسرے لوگ آپ سے ناجائز فائدہ اُٹھائیں گے۔	1	2	3	4	5
30	میں کام شروع کرنے سے پہلے کافی وقت ضائع کر لیتا ہوں/لیتی ہوں۔	1	2	3	4	5
31	میں شاذ و نادر ہی خوف یا پریشانی محسوس کرتا/ کرتی ہوں۔	1	2	3	4	5
32	میں اکثر اپنے آپ کو توانائی سے بھرپور محسوس کرتا/ کرتی ہوں۔	1	2	3	4	5
33	ماحول اور حالات کی وجہ سے پیدا ہونے والے موڈ یا احساسات کی جانب میری توجہ بہت کم ہوتی ہے۔	1	2	3	4	5
34	میرے جاننے والے اکثر لوگ مجھے پسند کرتے ہیں۔	1	2	3	4	5
35	میں اپنے مقاصد کی تکمیل کے لئے بہت محنت کرتا/ کرتی ہوں۔	1	2	3	4	5
36	لوگ میرے ساتھ جو سلوک کرتے ہیں، اس پر اکثر غصے ہو جاتا ہوں/جاتی ہوں۔	1	2	3	4	5
37	میں ایک خوش باش اور بلند حوصلہ شخص ہوں۔	1	2	3	4	5
38	میرا یقین ہے کہ ہمیں اخلاقی امور پر فیصلوں کے لئے مذہبی راہنماؤں سے رجوع کرنا چاہیے۔	1	2	3	4	5
39	کچھ لوگ مجھے سرد مہر اور خود غرض سمجھتے ہیں۔	1	2	3	4	5
40	جب میں کوئی منصوبہ شروع کروں، تو ہمیشہ اُسے ختم کر کے ہی دم لیتا ہوں/لیتی ہوں۔	1	2	3	4	5
41	اکثر اوقات جب کام خراب ہونے لگتا ہے، تو میں نا اُمید ہو کر اُسے چھوڑ دیتا ہوں/دیتی ہوں۔					
42	میں ایک زندہ دل اور روشن پہلو دیکھنے والا نہیں ہوں/والی نہیں ہوں۔	1	2	3	4	5
43	بعض اوقات شاعری کا مطالعہ کرتے ہوئے یا کوئی آرٹ کا شاہکار دیکھ کر میرے اندر سنسنی و جوش کی لہر دوڑتی ہے۔	1	2	3	4	5
44	میں اپنے رویوں میں سخت اور اڑیل ہوں۔	1	2	3	4	5
45	بعض اوقات میں اس حد تک قابلِ بھروسہ یا قابلِ اعتبار نہیں ہوتا/ہوتی، جس حد تک مجھے ہونا چاہیے۔	1	2	3	4	5
46	میں شاذ و نادر ہی اُداس یا غم زدہ ہوتا/ہوتی ہوں۔	1	2	3	4	5
47	میری زندگی میں تیز رفتاری نمایاں ہے۔	1	2	3	4	5
48	مجھے کائنات کے نظام یا انسانی حالت پہ غور و فکر کرنے میں کم دلچسپی ہے۔	1	2	3	4	5
49	میں عام طور پر دوسروں کی فکر اور خیال کرنے کی کوشش کرتا ہوں/ کرتی ہوں۔	1	2	3	4	5
50	میں ایک کارآمد شخص ہوں، جو ہمیشہ اپنا کام کر لیتا ہے۔	1	2	3	4	5

نمبر شمار	بیانات	مکمل غیر متفق	زیادہ تر غیر متفق	غیر جانبدار	زیادہ تر متفق	مکمل متفق
51	میں اکثر اپنے آپ کو بے بس محسوس کرتے ہوئے یہ چاہتا ہوں/چاہتی ہوں کہ کوئی اور میرے مسائل حل کر دے۔	1	2	3	4	5
52	میں نہایت سرگرم انسان ہوں۔	1	2	3	4	5
53	میرے اندر دانش ورانہ تجسس بہت زیادہ ہے۔	1	2	3	4	5
54	اگر لوگ مجھے ناپسند ہوں، تو میں انہیں بتا دیتا ہوں/بتا دیتی ہوں۔	1	2	3	4	5
55	مجھے لگتا ہے کہ میں خود کو کبھی منظم نہیں کر سکتا/سکتی۔	1	2	3	4	5
56	بعض اوقات شرم کے باعث چھپ جانے کو دل چاہتا ہے۔	1	2	3	4	5
57	میں دوسروں کا رہنما بننے کی بجائے اپنے طرز سے رہنا پسند کروں گا/کروں گی۔	1	2	3	4	5
58	میں اکثر اوقات نظریات اور تجریدی خیالات سے لطف اندوز ہوتا ہوں/ہوتی ہوں۔	1	2	3	4	5
59	اگر ضرورت پڑے، تو میں اپنے کام نکالنے کے لئے لوگوں کو استعمال کرنے پر آمادہ ہو جاتا ہوں/جاتی ہوں۔	1	2	3	4	5
60	میں ہر کام کو حدِ کمال تک کرنے کی کوشش کرتا ہوں/کرتی ہوں۔	1	2	3	4	5

Section-4

نوٹ: ذیل میں دیئے گئے سوالات کے دو ممکن جوابات ہیں

نمبر شمار	بیانات	ہاں	نہیں
1	کیا آپ نے پچھلے سال کوئی خیراتی کام کیا؟	1	2
2	کیا آپ کبھی کسی انتخابی نشست میں کھڑے ہوئے ہیں؟	1	2
3	کیا آپ نے کبھی کوئی رضا کارانہ کام کیا؟	1	2
4	کیا آپ نے قومی اسمبلی کے پچھلے الیکشن میں ووٹ دیا؟	1	2
5	کیا آپ نے کبھی خون دیا؟	1	2
6	کیا آپ باقائدگی سے مسجد جاتے ہیں؟	1	2
7	کیا آپ اس بات سے متفق ہیں۔ "لوگوں کے اوپر انحصار کیا جا سکتا ہے"؟	1	2
8	کیا آپ اس بات سے متفق ہیں۔ "قوانین توڑنے کے لیے بنائے جاتے ہیں"؟	1	2

A.3 Personalities Survey Instrument—Translation

Name

Designation

Union Council number

Name of BHU

HMIS code

Part 3

Medical Officer

(Self Reporting Section)

In this part of the on-going LUMS study, we are trying to collect data regarding the level of job satisfaction of health officers appointed in BHUs and the factors affecting their decision to retain their posts. We will be very thankful to you for taking some time out to fill out the form enclosed in this envelope, putting it back in and then handing it to the interviewer. We would like to remind you that, as with the rest of the survey, all of your responses for this section will be kept confidential by our research team and will not be shared by any official from the health department. Nevertheless, like before, your participation is voluntary.

Instructions for filling out the questionnaire:

1. Read every statement carefully and encircle the response you agree with.
    a. If you completely disagree with the statement, encircle (1).
    b. If you mostly disagree with the statement, encircle (2).
    c. If you are indifferent to the statement, encircle (3).
    d. If you mostly agree with the statement, encircle (4).
    e. If you completely agree with the statement, encircle (5).
2. This test has no concept of right or wrong, nor do you have to be an expert to solve it. Respond as sincerely as possible. Write your opinion as carefully and honestly as possible. Answer every question and ensure that for every response, you have encircled the right option. During the test, if you encircle the wrong option by mistake or if you change your mind after encircling a response, do not erase it. Instead, mark the wrong response with a cross and encircle your correct one.

Section 1

Statements:

1. Politics is a bad word
2. I respect elected officials who can convert good ideas to laws
3. The attitude of an elected official is just as important as his/her competency
4. I am indifferent to political give and take based on the concept of losing something to gain something
5. I don't care much for politicians
6. People do talk about the welfare of the general public but in reality they are only interested in their personal gains
7. It is very difficult for me to take a lot of interest in the events that take place in my community
8. I work selflessly for my community
9. Meaningful public service is really important to me
10. I would prefer that elected officials work for the welfare of the community even if it goes against my self interests
11. For a government employee, loyalty to the public should take precedence over loyalty to his/her officers
12. I consider serving the public my social responsibility
13. I believe that there are a lot of public issues that need to be addressed
14. I don't believe that the government can do anything to make the society more just
15. If any group is excluded from social welfare, we will stay in bad times
16. I am ready to spend every ounce of my energy to make this world a more just place
17. I am not afraid of raising my voice for the rights of others even if I am mocked for it
18. When government employees take their oaths, I believe that they are ready to take on responsibilities not expected from common citizens
19. I can go to any lengths to fulfill my civic responsibilities
20. Government service is the highest level of citizenship
21. I believe that no matter how busy a person is, it is his/her ethical responsibility to do his/her part in dealing with social issues
22. It is my responsibility to take care of the poor
23. The words 'work', 'honor' and 'country' evoke strong emotions in the bottom of my heart
24. It is my responsibility to solve the issues arising from mutual dependence of people
25. I am rarely moved by the plight of underprivileged people
26. A lot of social programs are very important and cannot be lived without
27. Whenever I see people in need, It becomes difficult for me to control my emotions
28. For me, working for the welfare of others is an expression of patriotism
29. I rarely think about the welfare of people I don't know personally
30. Day to day incidents make me appreciate time and again how much we depend on each other

31. I don't feel any sympathy for people who don't even bother to take the first step to fulfill their needs
32. There are only a few public programs that have my full support
33. For me, bringing a change in the society is more significant than personal success
34. I give obligations precedence over personal tasks
35. I consider being financially strong to be more important than doing good things
36. Most of the causes I work for are more important than my personal benefit
37. Serving the public is a source of satisfaction for me even if I don't get anything in return
38. I believe that people should give more to the society than what they take from it
39. I am one of the few people who are willing to help people even if it leads to personal losses
40. I am prepared for any sacrifice for the welfare of the society

Section 2

Statements:

1. I plan everything in advance
2. I take decisions quickly
3. I save routinely
4. When I am away from my work I am eager to go back to my work
5. I can think of a lot of occasions when I kept on working diligently while others gave up
6. I continue working on difficult projects even when others opposed it
7. I like working on multiple tasks at the same time
8. Rather than completing parts of multiple projects, I prefer to complete one project every day
9. I believe that it is better to complete old tasks before starting a new one
10. It is difficult to know who my real friends are
11. I don't try to do something that I'm not sure about
12. In general it can be said that the people in this area are honest and can be trusted
13. A person can become rich by taking risks
14. If, during the coming week, you inherit or receive a huge amount of money, would you still continue working with the health department?
15. How much money, if given to you, would convince you to leave your job or retire?
16. If someone finds your wallet which has Rs. 2000 in it, how likely do you expect is it that the wallet with the complete amount would be returned to you if the wallet was found by:
    a. Your neighbor
    b. The police
    c. A stranger

Section 3

Statements:

1. I am not depressed
2. I like to be amongst lots of people
3. I don't like to waste time day-dreaming
4. I try to be polite to everyone I meet
5. I keep all my things clean and tidy
6. I often feel inferior to other people
7. I laugh easily
8. When I find out the right way to do something, I stick with it
9. I often get into quarrels with my family members and coworkers
10. I pace my work such that I am able to complete everything on time
11. Sometimes when I am under intense psychological pressure, I feel as if I am about to fall to pieces
12. I don't consider myself to be a jolly person
13. Art and wonders of nature fascinate me
14. Some people think that I am selfish and egoistic
15. I am not a very organized person
16. I rarely feel lonely or sad
17. I really enjoy talking to people
18. I think that listening to controversial speakers can confuse students and lead them astray
19. I prefer cooperation over conflict
20. I try to complete all tasks entrusted to me according to my conscience
21. I often feel mentally stressed and anxious
22. I often long for thrilling situations
23. Poetry has very little or no influence on me
24. I am mistrustful and skeptical about the intentions of others
25. My objectives are very clear and I work to achieve them in a very organized way
26. Sometimes I feel completely worthless
27. I usually prefer to work alone
28. I often try new and exotic dishes
29. I believe that if you give them the chance, people will always exploit you
30. I waste a lot of time before starting to work
31. I rarely feel scared or depressed
32. I often feel full of energy
33. I don't pay much attention to the moods and feelings evoked my surroundings and circumstances
34. People who know me usually like me

35. I work very hard to achieve my goals
36. I often get frustrated by the way people treat me
37. I am a jolly and optimistic person
38. I believe that we should consult religious leaders for making decisions involving moral affairs
39. Some people think I am cold-hearted and selfish
40. When I start something, I don't rest until I finish it
41. Often when things start taking a turn for the worse, I give up and abandon my work
42. I am not a jolly and optimistic person
43. Sometimes while studying poetry or looking at masterpieces of art, I feel chills of thrill and excitement
44. I am strict and stubborn in my attitude
45. Sometimes I am not as trustworthy as I ought to be
46. I am rarely sad or depressed
47. Fast pace is a highlight of my life
48. I have little interest in pondering over the working of the universe or the human condition
49. I usually try to be concerned and care about others
50. I am useful person and always do my work
51. I often feel helpless and wish someone else would resolve my problems
52. I am a very active person
53. I have a lot of intellectual curiosity in me
54. If I don't like someone I let him/her know about it
55. I feel that I can never keep myself organized
56. Sometimes I want to hide myself due to shame
57. I would prefer to live on my own terms as opposed to being a leader for others
58. I often enjoy abstract ideas and theories
59. If need be, I am ready to use people to get my own work done
60. I try to do everything perfectly

Section 4

Note: The following questions have two possible answers

1. Did you do any charity work during the past year?
2. Have you ever contested for an electoral seat?
3. Have you ever done any volunteer work?
4. Did you vote in the last election for the National Assembly?
5. Have you ever donated blood?
6. Do you visit the Masjid regularly?
7. Do you agree with this statement: "People can be relied upon"

8. Do you agree with this statement: “Rules are made to be broken”

Protecting the Polls: The Effect of Observers on Election Fraud [1]

Joseph Asunka*, Sarah Brierley*, Miriam Golden*, Eric Kramon†, and George Ofosu*

*University of California, Los Angeles
†Stanford University

December 17, 2013

[1]Corresponding author: Eric Kramon, Center on Democracy, Development, and the Rule of Law, Stanford University; email: ekramon@stanford.edu. We received useful comments on earlier versions of this paper from audience members at the Annual Conference of The International Society for New Institutional Economics, June 20–22, 2013, Florence, Italy, at a Workshop on Electoral Integrity, Violence, and Vote-Buying, at the Center on Democracy, Development, and the Rule of Law, Program on Poverty and Governance, Stanford University, April 12–13, 2013 and at the American Political Science Association annual conference, Chicago, USA, August 29–September 1, 2013. Funding for the research reported here was provided by the U.K.'s Ghana office of the Department for International Development and by a National Science Foundation Grant for Rapid Response Research (RAPID), SES–1265247, neither of which bears any responsibility for the results reported here. We gratefully acknowledge the collaboration of our research partner in Ghana, the Centre for Democratic Development, as well as Ghana's Coalition of Domestic Election Observers. We also thank our 300 research assistants for data collection.

Abstract

Do domestic election observers deter electoral fraud? And under what conditions do political parties respond to the presence of observers to negate their impact? We address these questions by studying observers' effects on two markers of fraud — overvoting (more votes cast than registered voters) and unnaturally high levels of turnout — during Ghana's 2012 presidential elections. Our randomized saturation experimental design allows us to estimate observers' causal effects and to identify how political parties strategically respond to observers. We show that observers significantly reduce overvoting and suspicious turnout at polling stations to which they are deployed. We also find that political parties successfully relocate fraud from observed to unobserved stations in their historical strongholds, where they enjoy social penetration and political competition is low, whereas they are not able to do so in politically competitive constituencies. The findings have implications for understanding political party behavior and the effects of governance interventions.

1 Introduction

Elections are often marred by various types of malfeasance, including electoral fraud. According to the Database of Political Institutions (DPI)'s data, 20 percent of recent executive elections experienced so much fraud or intimidation that the outcome is affected (Keefer, 2002). An alternate dataset that identifies fraud only in cases where international election observers were present reports "moderate or major problems of election integrity" in a quarter of country-elections between 1980 and 2004.[1] Election fraud is common and it is often serious.

The deployment of election observers is one major response meant to enhance the integrity of elections in both developing and developed countries. Support for observer missions is a central aspect of democracy-promotion efforts by international and domestic actors in countries with new or fragile electoral institutions (Hyde, 2011; Kelley, 2012). Approximately 80 percent of elections that took place around the world in 2006 were monitored by observers. The main rationale for their deployment is that they are believed to be able to prevent or reduce electoral irregularities (Kelley, 2012). Nonetheless, little is known about the effectiveness of observers in reducing fraud, and the findings of various recent studies are fragmentary and difficult to compare.[2]

In this paper we address two questions. First, do observers reduce electoral fraud at the polling stations to which they are deployed? And second, under what conditions do political parties respond strategically to observers, thereby potentially negating their impact? We address these questions by studying the impact of domestic election observers on fraudulent activity during Ghana's December 2012 presidential and parliamentary elections. While Ghana is considered one of Africa's most stable and well functioning democracies, even there accusations of election fraud regularly mar the democratic process (e.g. Jockers, Kohnert and Nugent, 2010). These accusations may be politically consequential. They formed the heart of a legal challenge to the 2012

[1] Calculated from Kelley (2011).

[2] Observers may promote democratization through channels other than their direct effect on fraud. For example, they may also provide useful information about the conduct of elections (**?**).

presidential results brought by the country's major opposition political party on the grounds of widespread fraud and administrative irregularities. The Ghanaian Supreme Court held months of hearings broadcast live on television, until the justices decided in August 2013 that the incumbent president had been "validly elected."[3]

We partnered with Ghana's largest and most well established organization of domestic election observers, the Coalition of Domestic Elections Observers (CODEO), and randomly assigned election observers to just over $1,000$ of Ghana's $26,000$ polling places. Collecting data from an additional randomly selected $1,000$ polling stations to which observers were not deployed allows us to estimate the impact of observers on two indicators of election fraud: whether more votes are cast than voters are registered (*overvoting*), and whether abnormally high turnout rates are recorded.

The randomized saturation experimental design that we use allows us to measure the direct effects of observers and the strategic response of political actors to observer presence (Baird et al., 2012). Given the stakes in an election, those seeking to manipulate election results may respond to observer presence by relocating fraud to polling stations where observers are not stationed (Ichino and Schündeln, 2012). Where political actors respond strategically, random assignment of observers to polling stations (as in Enikolopov et al., 2013; Hyde, 2007, 2010; Sjoberg, 2012) is not sufficient to guarantee unbiased estimates of observers' causal effects because observers influence outcomes at both observed and unobserved stations (violating the Stable Unit Treatment Value Assumption). Our research design addresses this problem by randomly varying the percent of polling stations (saturation) with observers in a sample of electoral constituencies. We generate experimental estimates of spillover by comparing levels of fraud at unobserved stations at different levels of observer saturation. This information allows us to correct the estimates of observers' causal effects to incorporate spillover onto unobserved stations.

Measuring the strategic response of parties to electoral observation is important not simply because it allows us to accurately estimate the causal effects of observers. Analyzing variation in

[3]See `http://www.bbc.co.uk/news/world-africa-23878458`.

spillover across Ghana allows us to make inferences about an issue of broad theoretical importance that is difficult to study empirically: the question of the political and social conditions that facilitate the ability of political parties to coordinate election fraud (e.g. **?**Ziblatt, 2009). Empirical studies on this topic often rely on formal complaints lodged by contestants, legal petitions seeking to nullify election results (**??**Ziblatt, 2009), or reports from observers (Kelley, 2011), each of which may suffer reporting biases (Kelley, 2012; Lehoucq, 2003; Weidmann and Callen, 2013). The benefit of studying the spillover effects of observers is that, where we observe displacement, it suggests that the actors most interested in the election results — namely, major political parties — are capable of both orchestrating and of moving fraud within a matter of hours. Fraud displacement thus captures the coordination capacity of political parties to commit fraud.

Our results show that the net effect of observers is to reduce fraud. Observers reduce fraud at the stations where they are deployed by about 60 percent. We also find evidence that observers displace fraud to nearby but unobserved polling stations. This displacement is concentrated in the historical strongholds of Ghana's two major political parties. This suggests that parties are better able to relocate fraud in communities where they enjoy social penetration and where political competition is low. Finally, analyzing geographical information about the precise location of polling stations in one region produces more fine-grained results about deterrence and displacement. Observers deter fraud at the polling stations where they are stationed and at polling stations within one kilometer of those, whereas displacement takes place at ranges between one and five kilometers. We interpret these suggestive findings for what they tell us about the organizational capacities and activities of the political parties orchestrating fraud.

The remainder of this paper proceeds as follows. First, we provide a description of the setting of our work. Second, we present the hypotheses and detail the rationale for each. The third section presents the research design and sample selection strategy employed. We then turn to data and measurement and after that to our main findings. Section **??** examines the spatial dimension of

observers' spillover effects, including an analysis of geographically precise information from one region. We conclude with a discussion of some implications of our results.

2 The Setting

Ghana's December 2012 general election was the sixth following the country's return to democratic rule. Starting in 1992, Ghana has conducted competitive presidential and parliamentary elections every four years. Two of these elections (2000 and 2008) resulted in alternations of executive office. There are currently 275 Members of Parliament elected by plurality rule in single member constituencies.[4] The president is elected in a majoritarian run-off system. Our study took place during concurrent presidential and legislative elections, but is concerned with fraud only in the presidential contest. For the presidential contest the nation serves as a single constituency. As all votes count equally, candidates have incentives to seek votes everywhere.

Ghana has a stable two-party system, represented by the current governing National Democratic Congress (NDC) and the opposition New Patriotic Party (NPP). In the 2012 elections, the two parties together captured over 98 percent of the presidential vote. Both parties are multi-ethnic and multi-regional in composition (Whitfield, 2009) but each has regions where its support is particularly concentrated. In the four regions investigated in this study (Western, Central, Volta, and Ashanti), the NDC's stronghold areas include most constituencies in Volta whereas the NPP's stronghold comprises constituencies in the Ashanti region. These are the historic bastions of strength for the two main parties. Many of Ghana's political constituencies across the country's eight other regions remain moderately to highly competitive.

Despite a flawed transition election in 1992, Ghana's national elections since 1996 have usually been acclaimed as "free and fair" by both local and international observers. Nonetheless,

[4]Prior to Ghana's 2012 elections, the Electoral Commission by a Constitutional Instrument (CI 78) created 45 new constituencies in addition to the 230 constituencies used in the the country's 2004 and 2008 presidential elections. See `http://politics.myjoyonline.com/pages/news/201209/94714.php`.

elections have not been devoid of allegations of fraud and electoral malpractice (Jockers, Kohnert and Nugent, 2010), though the magnitude of irregularities and their impact on electoral outcomes remains unknown. In almost all elections, political parties have alleged fraud before, during, and after the polls. Allegations of electoral irregularities in elections before December 2012 frequently claimed the existence of ghost voters on the electoral register, illegal voting by minors and foreign nationals — allegedly from across the border in Togo — intimidation of voters and party agents by national security forces and political parties, ballot stuffing, and tampering with results during transmission from polling stations to collation centers (Smith, 2002). Jockers, Kohnert and Nugent (2010) note suspiciously high turnout rates at some polling stations in 2008, especially in the stronghold regions of the two major political parties, Ashanti and Volta.

In 2012, biometric voter registration and polling-place biometric verification processes were introduced in an attempt to eliminate irregularities that had occurred in previous elections. The Electoral Commission adopted a policy that only persons whose identities were verified by the biometric verification machine at the polling station would be permitted to exercise the franchise. However, the verification machines broke down in approximately 19 percent of polling stations.[5] By noon on election day, the president had appealed to the Electoral Commission to consider allowing people with valid voter ID cards to vote at polling stations where the biometric verification machines were not functioning.[6] This reintroduced the possibility of double-voting, impersonation, and the other irregularities that the biometric process was intended to eliminate.

Presidential elections in Ghana have become extremely competitive. The 2008 presidential election was won by 40,000 votes out of an electorate of roughly 14 million, a small enough margin that even relatively modest levels of fraud might have affected the outcome. In the 2012 elections, the NDC candidate, John Dramani Mahama, was declared winner of the presidency with 50.7 percent of the vote and a margin of just over $300,000$ votes. Turnout increased from 70 percent

[5]This figure is calculated using data from the 4,000 CODEO observers dispatched across the country on election day.

[6]See `http://politics.myjoyonline.com/pages/news/201212/98391.php`.

in the prior presidential race to 80 percent after intense mobilization efforts by the two major parties. Despite the NDC's larger margin of victory, the NPP subsequently contested the outcome in the Supreme Court, citing instances of fraud and electoral malpractice. The NPP's legal petition focused on overvoting, which it defined as more votes recorded on the summary of results reported by the polling station than ballots issued or than voters registered at the station. The NPP asserted that overvoting had occurred in more than $1,800$ polling stations in Ghana and that turnout had exceeded 100 percent in more than 60.

As our study took place in four of Ghana's ten regions, which were not selected to be nationally representative, our data cannot be used to evaluate the validity of the NPP's petition. Our data does, however, corroborate that fraud did occur. Combining official data from the Electoral Commission with data that we collected from a sample of approximately 2,000 polling stations, we document instances of either overvoting — more voters casting votes in the presidential election than were officially registered — or ballot stuffing — more ballots found in the presidential ballot box than were known to be cast — in about 8.5 percent of polling stations. Turnout rates in some polling stations were also unusually high, sometimes close to and over 100 percent, as we describe in more detail below.

2.1 Domestic Election Observers in Ghana

Domestic election observation has a well-established history in Ghanaian elections. To combat electoral malpractice and enhance the credibility of the electoral process, a coalition of civic organizations came together to observe the 1996 general elections. In the 2000 elections, election observation was formalized with the formation of the Coalition of Domestic Election Observers (CODEO). This organization has led domestic election observation since the 2000 elections. CODEO is widely seen by Ghanaians as non-partisan and independent, and there is broad acceptance of their role in the nation's political life (Boafo-Arthur, 2006, p. 53).

All observers receive formal training, which includes swearing a public oath that they will act impartially and support the conduct of free and fair elections. Observers are also accredited by Ghana's Electoral Commission. Once accredited, observers are granted the right to access and observe proceedings at any polling station or collation center. Most CODEO observers are posted to a single station that they observe from the opening to the close of the polls, including the public vote count that takes place at the end of the day at each polling place. CODEO observers wear uniforms that identify their position and affiliation. A typical polling stations is set up with a desk for the presiding officer who oversees the operation of the station. He or she is assisted by several polling officials who verify voters, and tear and stamp ballot slips (which verifies their authenticity). There is a seating area for political party agents, and two voting areas where voters fingerprint their ballot slips behind cardboard screens, one for the presidential and the other for the parliamentary election. The observers usually position themselves away from other officials. In 2012 all observers were trained to use SMS to report polling station activities, including any irregularities and disruptions to a national data center. If an incident is serious, CODEO has communication structures in place to alert appropriate legal and security officials. CODEO also releases press statements throughout the day so the public has immediate access to information on the progress of the election.

3 Hypotheses

Electoral fraud comprises a complex set of activities that may occur at many points in the electoral process. Before election day, fraud may affect voter registration (Ichino and Schündeln, 2012; Hidalgo and Nichter, 2013). On election day and perhaps later, it may occur when votes are aggregated at regional or national tabulation centers. Our focus is on objective measures of polling station level fraud. Data on the number of people who vote at a station and the number of ballots in the ballot box, as well as turnout, are objective measures that are easy to collect. This is also the level of fraud which observers are deployed to prevent. Our focus is therefore narrower than

that of Lehoucq (2003), for instance, which offers a broad classification of fraud that comprises all "clandestine efforts to shape election results" (p. 233).

We expect fraud to be a function of its potential return to election contestants as well as the cost of execution. With respect to the former, the highly competitive nature of presidential elections in Ghana provides substantial motivation for fraud. Polls conducted prior to the December 2012 election suggested that the outcome was too close to call.[7] In addition, Ghana's electoral rules dictate a runoff election should none of the presidential contenders receive over 50 percent of the vote. Because both of Ghana's two major parties enjoy support from roughly equal proportions of voters, small shifts in the votes received by each of their presidential candidates as well as small increases or decreases in turnout could push one of the main contenders above or below the 50 percent threshold. In this context, there are high potential returns to using fraud to increase votes.

While the presence of an election observer should have no effect on the potential electoral returns to fraud, observers influence the costs to political parties of engaging in fraud. These costs could be legal, if observers report instances of fraud to authorities, or reputational, if observer reports are publicized by the media or to the local community. They could also be logistical: the costs of attempting to hide fraud will be higher when observers are present. For these reasons, we expect that *observers will reduce fraudulent activities at polling places where they are stationed (H1).*

Political parties in Ghana are among the most institutionalized and organized in Africa. Both major parties are organized hierarchically and elect party representatives at the polling station, constituency, regional and national level. The NDC and NPP maintain party offices across the country both during and between elections. Both parties attract large numbers of activists (**?**). The strong level of party organization leads us to expect that political parties have the potential to respond to the presence of an election observer in the polling place by relocating efforts of fraud.

[7] Reported in `http://www.ghanaweb.com/GhanaHomePage/NewsArchive/artikel.php?ID=250900`.

We therefore expect that *the presence of an observer will displace fraudulent behavior to nearby polling places (H2).*

We are agnostic about where displacement will occur. It could be that *displacement effects will be more concentrated in urban than rural areas (H3a).* The rationale is that transportation networks are more developed and polling stations closer together in urban localities. These factors allow parties to more efficiently relocate fraud. Conversely, rural polling stations may be more socially homogeneous, where voters may be more complicit or parties may more easily co-opt polling station officials. Also, rural polling stations may be under less scrutiny from journalists or unofficial observers. It could therefore be hypothesized that *displacement effects will be more concentrated in rural than urban areas (H3b).*

Finally, fraud is also facilitated by the local resources available to party organizations. Here we distinguish between areas of the country that are party strongholds and areas in which neither party is electorally dominant. In stronghold regions party agents and electoral officials might be more easily corrupted by peer pressure and the social expectation to support the dominant party. They might also feel more pressure because of their own personal links or financial reliance on the local political networks that the dominant party controls (as in Ziblatt, 2009). There is also a greater probability that these officials will themselves be avid party supporters. The same considerations apply to ordinary citizens, who may therefore be more willing to facilitate or, at the very least, not denounce fraud. Finally, the dominant political party will control more resources on the ground and be capable of mobilizing more activists and sympathizers. We thus hypothesize that *observers will have a greater impact on reducing fraudulent activities in party strongholds than in competitive constituencies (H4).* On the other hand, the stronger local party networks that characterize party strongholds also imply that displacement effects could be substantially larger in these environments, where the dominant party may find it easier to relocate fraud to a neighboring polling station to avoid an observer. Our final hypothesis is thus that *the effect of observers on the*

displacement of fraudulent activities to nearby polling places will be greater in party strongholds (H5).

4 Research Design

Observational approaches to studying observer effects (as in Herron (2010)) are likely to generate biased estimates of observers' impacts. Observers may for example be deployed to polling places where fraud or irregularities are expected to occur precisely in order to prevent such activities. If this is the case, fraud could be more prevalent in observed polling places even if observers reduced fraud where they were present. Observers may also be deployed to geographically convenient locations that differ in important and potentially difficult to measure respects from stations without observers. In each of these instances, omitted variables could bias estimates of observer impact.

To address this selection problem, a number of recent studies have integrated randomization or pseudo-randomization into their research designs. Exploiting an opportunity in which international election observers were assigned "almost as if" randomly in Armenia, Hyde (2007) finds that they reduced the vote share received by the incumbent, who had been perceived as likely to steal the election. Hyde (2010) randomly assigns locations to a small group of international observers during an election in Indonesia and reports that observers increase the vote share of the incumbent, an unexpected (and not easily explicable) outcome. Sjoberg (2012) studies the impact of random assignments of domestic election observers in Azerbaijan, Georgia, and Kyrgyzstan and finds that observers reduce unreasonably high levels of turnout and have mixed effects on fraud at the vote counting stage. Enikolopov et al. (2013) randomly assigns observers to polling stations in Moscow during Russia's 2011 parliamentary elections and shows that observers substantially reduce the vote share of the incumbent in a context in which the incumbent party is reputed to be responsible for fraudulently inflating its vote share.

The potential for political actors to respond strategically to observer presence implies that randomization alone is not sufficient to guarantee unbiased estimates of observers' causal effects. Observers may also impact outcomes at unobserved (control) polling stations, either by inducing the strategic relocation of fraud to unobserved stations (displacement) or by deterring fraud at these stations. Ichino and Schündeln's (2012) study of the voter registration process in Ghana provides evidence of the former, illustrating that observer presence led to the relocation of fraudulent activity to nearby voter registration centers without observers.[8] In Moscow, Enikolopov et al. (2013) report evidence of the latter: the presence of observers at one polling station in an apartment building reduces fraud at the second (unobserved) polling station in the same building. In both studies, observers influence outcomes at control polling places, which implies a violation of the Stable Unit Treatment Value Assumption (SUTVA); a requirement for mean comparisons between treatment and control groups to yield unbiased estimates of causal effects (**?**). To address the challenge posed by the potential for both omitted variables bias and spillover, we implement a randomized saturation experimental design (Baird et al., 2012), which we discuss below.

4.1 Sample Selection and Treatment Assignment

In the 2012 general elections, CODEO deployed observers to over $4,000$ of Ghana's $26,000$ polling places. We worked with CODEO to randomize the placement of $1,000$ of their observers in four regions in southern Ghana: Volta Region, the longstanding stronghold of the ruling NDC; Ashanti, the stronghold of the main opposition NPP; and Western and Central, which represent electorally competitive regions.[9] These four regions therefore vary in terms of electoral competitiveness and

[8] The presence of election observers may also increase electoral manipulation of other types, such as in pre-election public spending (Hyde and O'Mahony, 2010), vote buying, violence, intimidation, and so on (Hyde, 2011; Kelley, 2012; Simpser and Donno, 2012). Although important, these kind of displacement effects are not the focus of this paper.

[9] We excluded the region of Greater Accra, where Ghana's capital is located, because we anticipated that international election observers might focus on these easy-to-reach polling stations which could contaminate the treatment.

the social penetration of political parties, which we hypothesize is theoretically important for the displacement of fraud.

We randomly selected 60 constituencies, blocking on the level of electoral competition and polling station density of each constituency. We define as competitive those constituencies in which the margin of victory in the 2008 presidential elections was not more than 10 percentage points, which in Ghana is easily reversible.[10] Polling station density, which we use as a proxy to distinguish rural from more urban areas, represents the approximate number of polling places per square kilometer in each constituency. We define as high density those constituencies with a number higher than the median (seven polling places per square kilometer). We weight the number of constituencies selected from each region by a measure of the regional population.[11] We randomly selected 30 percent of the polling stations in each of these 60 constituencies to form our sample.

To implement the randomized saturation design, we follow a two-step procedure for assigning treatment (domestic election observers) to individual polling stations. First, we randomly assign to our sample of constituencies one of three constituency-level observer saturations: 30 percent, 50 percent, or 80 percent of polling stations in the sample.[12] Second, we randomly assign observers to polling places within constituencies according to the probability dictated by the constituency's saturation. In assigning the constituency level saturation level, we stratify on electoral competitiveness and polling station density.

Figure 1 illustrates the research design in the Ashanti region. Each sampled polling station is depicted on the map (constituencies with no polling stations in the sample are so marked). Observed stations are indicated by circles and unobserved stations by triangles. There is evident

[10]The average margin overturned in swing constituencies in 2008 was 12.45 percentage points.

[11]The number of constituencies in each region reflects the region's population by construction. We selected 23 constituencies from Ashanti region, the most populous in the country; 11 from Central Region; 13 from Volta; and 13 from Western.

[12]This means that roughly 5 percent, 15 percent, and 30 percent of polling stations in the entire constituency are selected for observation as our initial sample is 30 percent of all polling station in the constituency.

variation across constituencies in the share of sampled polling stations that were assigned an observer.

Insert Figure 1 about here

Random assignment of observer saturations allows us to generate experimental estimates of spillover within constituencies. We generate these estimates by comparing outcomes in control polling stations at each of the three observer saturation levels. If fraudulent activity increases in control polling stations as observer saturation increases, this is evidence of a displacement effect. On the other hand, if fraudulent activity decreases in control polling stations as observer saturation increases, this is evidence of a deterrent effect. In ways described in detail below, we use these estimates to adjust our estimates of observers' causal effects to correct for the bias generated by spillover. Varying the saturation of observers at the constituency level also allows us to measure the net effect of observers on fraudulent activity within a constituency, accounting for the potential for spillover effects. Thus, the strength of this design is that it permits causal statements about observers' direct effects, about spillover effects, about the equilibrium effects of observers when spillovers are taken into consideration, and about the saturation threshold that reduces or eradicates fraud entirely, if such a threshold exists.

5 Measuring Fraud

Measuring fraud directly poses a challenge since it is generally conducted clandestinely. In this paper, we analyze two measures of potential polling station fraud on election day.[13] Our measures rely on objective information about polling place election results and voter turnout gathered from over $2,000$ observed and unobserved polling places after voting was completed on election day.

[13] In addition to results for the two measures reported here, we also collected information on ballot stuffing, defined as more ballots in the presidential ballot box than voters known to have cast votes. The measure was created by asking polling station officials whether more ballots were found in the box than voters known to have voted. Results for ballot stuffing are reported in Appendix **??**. Although the pattern of results for ballot stuffing is consistent with those reported for overvoting and turnout, we put them in an appendix for reasons of space and to ease presentation.

In Ghana, law requires that votes be counted in public at each polling station after the polls close, making it possible to gather polling station level information before it is aggregated (and potentially tampered with) at higher levels (Public Elections Regulations, 2012 (C.I.75)).

Our first measure identifies whether more voters cast votes in the presidential election at a polling station than were officially registered to do so. Voters are only legally allowed to vote at the polling station where they are registered. This measure, which we call *overvoting*, is a marker of potential fraud since it suggests that unregistered voters cast ballots, that double voting occurred, or that ballot stuffing took place. To measure overvoting, we obtain official figures from the Electoral Commission on the number of registered voters at each polling station. These registration figures were released before election day. Observers and trained enumerators collected data on the number of votes cast in our sample polling places, as reported on the official polling station results forms. We create a dichotomous measure of overvoting that takes a value of 1 if a polling place had more people vote than the number of voters registered.

The limitation of the dichotomous overvoting measure is that it does not include information on the amount of overvoting that occurred, only the frequency of polling stations affected. Moreover, it codes polling places with unreasonably high turnout rates — such as those with 98 or 99 percent turnout — identically to those with turnout rates closer to the national average. We therefore follow previous literature and proxy for electoral fraud using the polling station turnout rate as our second outcome variable.[14] If turnout correlates negatively with the placement of observers, this is evidence of potential fraud because it shows that turnout is inflated where observers are not available to monitor the election process.

Table 1 presents descriptive information about the two main outcome variables. We report means with standard deviations in parentheses. The average turnout rate among polling stations in

[14] In the planning stages of this project, we hypothesized that observers might have the effect of increasing polling station turnout. Our reasoning was that observers might reduce violence or the intimidation of voters, thereby subsequently leading to increased participation. Ghana's 2012 election was not marked by violence, and we now believe that turnout is a better indicator for fraud than for enhanced voter participation.

our sample is 83 percent, close to the national turnout rate of 80 percent. Columns 2 and 3 provide preliminary information on the effects of observers. At observed polling places, the overvoting rate is 2.6 percent, while among unobserved polling places it is 7 percent. The average turnout rate is 82 percent at observed polling places, while it is 86 percent at unobserved. These differences, all of which are statistically significant at the 0.05 level using two-tailed difference of proportions tests, provide preliminary evidence of fraud by showing systematically better outcomes on both fraud indicators where election observers are present. The remainder of the columns show how the rates vary across different types of constituencies, and are discussed later in the paper.

Insert Table 1 about here

6 Covariate Balance Tests

Before turning to our results, we first demonstrate that the communities in which our observed and unobserved polling places are located are comparable across a range of political and socio-economic pre-election covariates. We use data from a household survey we conducted in the communities near observed and unobserved polling places during the two days following the elections.[15] As part of the survey, we gathered data on voting behavior in the prior 2008 election as well as measures of socio-economic conditions.

Table 2 presents means in control and observed communities on a number of pre-election covariates. It also presents the difference in these means and the p-value of a two-tailed difference-of-means test. The first section of the table shows that the partisan voting histories of residents near observed and unobserved polling are comparable. In both sets of communities, about 35 percent report voting for the NPP in the 2008 presidential election, while about 43 percent report voting for the NDC, whose candidate was the winner of that election. The remaining sections of the table

[15] We surveyed over 6,000 Ghanaians. Ideally, we would have randomly sampled individuals from the official voter register. As this was not available, we instead employed the random sampling techniques used across Africa by the Afrobarometer public opinion survey. Our enumerators visited each sampled polling place and then selected four households using a random walk technique.

examine measures of education, poverty and well-being. Observed and control polling stations are also similar along these dimensions. The data presented in the table shows that the communities surrounding our observed and control polling stations are comparable across a range of political and socio-economic characteristics that may be thought to affect the level of fraud.

Table 2 about here

7 Direct and Spillover Effects of Observers

7.1 Direct Observer Effects on Indicators of Fraud

Do domestic election observers reduce fraudulent activity in the polling places to which they are deployed? We first estimate the direct effect of the presence of an observer on each of our two indicators of fraud — overvoting and turnout. All models are estimated using OLS.[16] As saturation is assigned at the constituency level, we report robust standard errors clustered by constituency.

Table 3 presents the results. Column 1 reports the direct treatment effect of observers on overvoting in a model without covariates. Column 2 introduces controls for each of our blocking variables (competition and polling station density) and constituency-level observer saturation indicators. Consistent with our first hypothesis, we find that observers reduce the probability of overvoting at a polling station by 4.5 percentage points. This effect corresponds to a roughly 60 percent reduction in the probability of overvoting from the control group mean of 7 percent. The results in columns 3 and 4 show that the presence of an observer also reduces the polling station turnout rate. We find a significant and negative observer effect: the turnout rate where observers are present is between 5 and 5.5 percentage points lower than where they are not. These results provide initial evidence that observers are effective in reducing fraudulent activity in the polling places where they are stationed.

Table 3 about here.

[16] Overvoting inferences are robust to estimations that use logistic regressions. We report OLS results for ease of interpretation.

These estimates may be biased by observers' spillover effects onto unobserved polling stations. An unbiased estimate of observers' direct effects would compare outcomes in treated polling places to those in a "pure control" group. In this case, a pure control would be an unobserved polling station located in a constituency without any observers. Although ideal for estimation purposes, we were unable to create such control constituencies. The mission of our partner, Ghana's electoral observation group, is to reduce fraud and improve the overall quality of elections. There were therefore programmatic reasons to send observers to every constituency in the country.

In the absence of a pure control group, we use our data to construct a set of estimated pure control outcomes. We use these estimated pure control outcomes to adjust our causal effect estimates. We adopt two main approaches. In our first approach, we use outcomes in control stations in the low saturation constituencies as our estimate of the pure control outcome. In these constituencies, fewer than 10 percent of all polling stations are observed, which ensures minimal contamination from spillover. In our second approach, we assume a linear relationship between constituency observer saturation and outcomes in the control group (**?**). This functional form assumption allows us to generate a set of predicted values in the control group where saturation is equal to zero. The advantage of the first approach is that it uses the actual data and requires no functional form assumption. The downside, however, is that the estimate of the pure control group average may be biased because of spillover. The second approach, on the other hand, is not susceptible to bias from spillover but can be sensitive to the assumption of a linear saturation effect.

In the second approach, the estimation process is as follows. We first estimate the naive (direct) treatment effect (θ) of observers:

$$Y_{ij} = \alpha + \theta T_{ij} + \varepsilon_j \tag{1}$$

where Y is a measure of fraud, T is treatment with an observer, subscripts i and j are polling station and constituency markers, and ε is an error term clustered at the constituency level. In this equation, which corresponds to the models in columns 1 and 3 of Table 3, α represents the average in the control group.

We then assume a linear saturation effect and estimate the following model:

$$Y_{ij} = \alpha_0 + \theta_0 T_{ij} + \lambda S_j + \gamma T_{ij} * S_j + \varepsilon_j \tag{2}$$

In this equation, T_{ij} is the observer treatment indicator, S_j is a continuous measure of constituency observer saturation, and $T_{ij} * S_j$ is the interaction between the two. α_0 therefore represents an estimate of fraud in control stations in constituencies where saturation is equal to zero — the pure control outcome. The estimated average bias in the control group is the difference between the observed average in the control group (α) and the estimated pure control outcome, $\alpha - \alpha_0$. The adjusted treatment effect is therefore given by $\theta + (\alpha - \alpha_0)$ (**?**).

Our adjusted estimates, presented in Table 4, do not change our finding that observers reduce fraudulent activity at the stations to which they are deployed. Panel A presents results on overvoting. We find evidence of a relatively small upward bias in our estimate of the direct observer effect. Adjusting the treatment effect using controls from the lowest saturation constituencies generates an estimate that observers reduce fraud by 4.4 percentage points (0.1 percentage points less than the unadjusted estimate). With the linearization approach, we estimate that observers reduce polling stations affected by overvoting by 3.7 percentage points. The true effect is likely to lie somewhere between these two estimates. With respect to voter turnout (Panel B), using controls from low saturation constituencies we find that the unadjusted estimate is slightly biased downwards; the data show that observers reduce fraud by 5.5 percentage points (0.3 percent more than the unadjusted estimate). The linear estimate suggests however that observers' effects on fraud are half what the unadjusted estimate predicts, with observers reducing turnout by 2.5 percentage

points. Overall, the adjusted effects support the claim that observers reduce fraud: they reduce the probability of overvoting by roughly 4 percent and turnout by somewhere between 2.5 and 5.5 percent.

Table 4 about here.

In additional analyses (results not shown), we find mixed support for the hypothesis that observers have a greater direct effect of reducing fraud in party strongholds ($H4$). While we find a larger effect of observers on overvoting in strongholds using both of our adjustment methods, we find inconsistent support for turnout.

7.2 Spillover Effects within Constituencies

Under what conditions are political parties able to respond strategically to the presence of observers and relocate fraud? Our randomized saturation design generates experimental estimates of observers' spillover effects, addressing this question. To identify spillover, we compare outcomes in control polling places at each level of observer saturation. An increase in fraudulent activity in control polling places as observer saturation increases is evidence that observers displace fraud to polling places within the same constituency. A decrease in fraud in control polling places as the observer saturation increases is evidence of a deterrent effect.

We hypothesized that observer displacement effects might vary with the partisan environment, and in particular that spillover might be more intense in the strongholds of Ghana's two major political parties. In these constituencies we hypothesize that parties will find it easier to shift fraud due to the higher concentration of political supporters and partisan networks, and greater potential to co-opt election officials to conduct fraudulent activities. The results support this hypothesis.

In these analyses, we define political party strongholds more narrowly than in our initial blocking variable. The results in this section are based on a definition of party strongholds as constituencies where the same political party has won over 65 percent of the vote in each of the four presidential elections held starting in 1996 (1996, 2000, 2004, 2008). While the results are

comparable if we use the initial blocking variable, we use this measure of a party stronghold because it restricts the analysis to those constituencies where we believe parties are most likely to have the social penetration required to coordinate electoral fraud.[17]

In these analyses, we make our measure of constituency observer saturation continuous and estimate a model that includes a triple interaction between observer presence, observer saturation, and party stronghold status. As the triple interaction is difficult to interpret, we present the main results of interest graphically. The regression results are presented in the Appendix A.

Figure 2 plots the relation between observer saturation and overvoting rates for *unobserved* (control) polling stations. In political party strongholds, the overvoting rate in the control group increases from 7 percent in low saturation areas to over 10 percent in high saturation areas. In contrast, in the competitive constituencies the overvoting rate decreases slightly as the saturation increases. Figure 3 provides evidence of similar spillover effects with respect to voter turnout. The control group turnout rate increases substantially across the saturation distribution in political party strongholds, while the opposite pattern holds in competitive constituencies.

Figure 2 about here.

Figure 3 about here.

With respect to overvoting and voter turnout, we therefore find evidence of a displacement effect in party strongholds and a mild deterrent effect in electorally competitive areas. These results illustrate that observers do have spillover effects onto unobserved polling stations, consistent with $H2$, but that the nature of the spillover is conditional on the local political environment to which observers are deployed. Where parties are dominant, they are able to coordinate fraud to unobserved stations, corroborating our fifth hypothesis. In more competitive settings, observers deter fraud in unobserved stations. In additional analyses (results not presented), we examine whether spillover effects are different in urban and rural areas ($H3$) and do not find evidence that they are.

[17] In the initial blocking variable we define as competitive constituencies in which the margin of victory in the 2008 presidential elections was 10 percentage points or less. Using this definition, 38 constituencies of our total sample of 60 were defined as low competition and the remaining 22 as high competition.

8 Spatial Aspects of Spillover

The previous section provides evidence of general spillover effects within constituencies. Observers reduce fraud in polling stations where they are stationed but push some of it into polling stations in the same constituency. In this section, we investigate the spatial aspect of this spillover. Analyzing spatial spillover allows us to derive more precise conclusions about the movement of fraud and to better understand the ways in which parties coordinate fraud within constituencies. Our results highlight two main patterns. First, the general spillover effects documented above are largely made up of spatial spillover from observed stations to geographically proximate stations without observers. Second, spatial spillover may vary with distance: we find suggestive evidence that observers have a deterrent effect on control stations within one kilometer of treated stations and that spillover occurs at distances between one and five kilometers.

We estimate spatial spillover effects in two ways. First, we consider spatial spillovers within Electoral Areas (EAs). EAs are political units smaller than constituencies from which voters elect local government councilors (we provide more details below). Second, we analyze spatial effects by geographic distance within constituency boundaries. With respect to estimation, our randomization process places stations into four potential experimental conditions:

1. Y_{01}– treated stations with no other observers present in the same EA or within d kilometers
2. Y_{11}– treated stations with at least one other observer in the same EA or within d kilometers
3. Y_{10}– control stations with at least one observer present in the same EA or within d kilometers
4. Y_{00}– control stations with no observers in the same EA or within d kilometers

To calculate spatial spillover effects on control stations, we compare outcomes in Y_{10} to outcomes in Y_{00}. That is, we examine whether outcomes in control polling stations with an observer nearby are different from those at control polling stations without any observers nearby. To

calculate spatial spillover effects on observed stations, we compare outcomes in Y_{11} to outcomes in Y_{01}. That is, we examine whether outcomes in observed polling stations with another observer nearby are different from those at observed polling places without another observer nearby.

Conditional on constituency saturation, each polling station has an equal probability of being treated with an observer. However, polling stations are not distributed equally in space or across EAs. As a result, each sampled polling station does not have an equal probability of landing in each of the four experimental conditions. This difference in treatment assignment probability implies that a simple difference-in-means comparison between the relevant experimental groups would generate biased estimates of spatial spillover effects (Gerber and Green, 2012, pg. 270).

We therefore estimate treatment assignment probabilities using simulation methods. We replicate our treatment assignment process $10,000$ times and use the distribution of treatment assignments to calculate the probabilities. We then estimate the difference in *weighted* means between the relevant experimental groups, where we weight units by the inverse probability that they are assigned to their actual condition (Gerber and Green, 2012). To characterize the uncertainty of our estimates, we use randomization inference assuming constant effects across all units.[18]

8.1 Spatial Spillover Within Local Political Units

Each parliamentary constituency in Ghana is divided into approximately 25 Electoral Areas, with approximately 5 polling stations within each. Polling places in the same EA are closer to one another than they are to most other polling places outside the EA.[19] We can therefore use EAs to study the effect on markers of fraud of having an observer nearby. Polling stations in our sample are distributed roughly evenly across each of the four spatial experimental conditions enumerated above: Y_{01} (N=574), Y_{11} (N=718), Y_{10} (N=567), and Y_{00} (N=471).

[18] See (Gerber and Green, 2012, ch. 2) for details.

[19] This may not be the case for polling places on the border of two EAs, but because we do not have GIS locations of polling places in all four regions studied, in this section we proxy polling place proximity using EA designation.

Table 5 presents estimated spillover effects and 95 percent confidence intervals. We find evidence of spatial displacement in the full sample of constituencies. Polling stations without an observer are 6.6 percentage points more likely to experience overvoting if there is another observer present in the same EA. Similarly for turnout, unobserved stations have a turnout rate 7.4 percentage points higher when there is an observer in the same EA. We also find evidence that these spatial displacement effects are larger and mostly concentrated in stronghold constituencies. The within-constituency spillover documented in Section **??** appears to be driven by local spatial spillover patterns. Additionally, results show that all of the spillover effects on observed stations are substantively small and none are statistically distinguishable from 0. This implies that the presence of an observer protects a polling station from the potentially negative impact of spillover from nearby stations.

Table 5 about here.

8.2 Spatial Spillover across Different Distances

In this section, we use information about the precise location of polling places in the most populous of our study regions, the Ashanti, to provide more spatially precise information about spillover effects. We use data from 817 polling stations in the region, 52 percent (421) of which were observed and 48 percent (396) of which were unobserved. We refer readers back to Figure 1 for a depiction of the spatial spread of the stations. We use GIS data from all of the 23 constituencies we sample in the region.[20]

We calculate the distance from each polling station to all other stations in the same constituency. As in our randomized saturation design, we assume that spillovers occur within constituencies and do not cross constituency boundaries.[21] Within constituencies, the average distance between one sampled polling station and other polling stations is 11 kilometers, with a standard

[20]This data collection took place in July and August 2013. Resource constraints allowed us to collect geo-coded polling station data from only one region. We choose the most populous region in the sample.

[21]The justification for this is that political parties are organized on a constituency-basis in Ghana.

deviation of 2 kilometers. This average ranges from 1.3 to 27 kilometers across constituencies. We estimate spatial displacement effects by calculating weighted differences in means and use randomization inference to estimate confidence bounds for these estimates at the 80 percent confidence level.

Figure 4 displays the estimated spatial spillover effects. The relatively small sample size in each experimental group means that the findings presented in this section are only suggestive. The results show that the spillover effects of observers appear to vary depending on proximity. The first panel in the Figure 4 shows that observers have a deterrent effect on polling stations that fall within a one kilometer radius. The probability of overvoting in control stations falls by 5 percentage points. At the same time, observers increase the probability of overvoting at unobserved stations located within a two to five kilometer radius by 2.6 percentage points and 2.3 percentage points, respectively. With respect to overvoting, deterrence thus shifts to displacement as proximity to an observed station grows. In contrast, we find that observers increase the level of turnout in unobserved stations regardless of distance. The magnitude of spillover is higher in stations that are further away (i.e. five kilometers) from observed stations compared to those close (i.e. one and two kilometers). While turnout increases by about 4 percentage points in short distances, they increase by more than 6 percentage points in polling stations further than five kilometers from observed polling stations.

Figure 4 about here.

The results presented in this section show that the spillover effects of election observers on fraud are spatially differentiated within constituency boundaries. They also offer insights into how political parties operate to coordinate fraud. That displacement effects are spatially concentrated suggests that the coordination of fraud within constituencies may not be centralized. If it were, we would find that the probability of observing fraud in unobserved polling stations within the same constituency would be roughly similar regardless of distance from observed stations. Instead, the spatial patterns that we uncover suggest that party operatives intending to commit fraud likely

operate on the ground in a decentralized manner, communicating with others known personally to them within their geographically proximate party networks.

9 Do Observers Reduce Overall Rates of Fraudulent Activity in Constituencies?

The findings presented thus far raise an additional and important question about the net effects of observers. Do observers simply displace fraud, thereby generating no overall reduction, or does their presence at some polling stations reduce the aggregate level of fraud? To answer this question, we aggregate the data up to the level of our 60 constituencies and generate constituency-level rates of overvoting and average rates of voter turnout. We then examine how these constituency-level measures vary with the share of observed polling places in the constituency.

Table 6 presents the results. In the first three columns, the dependent variable is the share of polling places in each constituency that exhibit evidence of overvoting. Each model is estimated using ordinary least squares. Column 1 presents results from the full sample of constituencies. We find that increasing the share of observed polling places in a constituency results in a net reduction in the rate of overvoting. If we increase observer saturation by 50 percent, the equivalent of moving from our low to high saturation treatments, rates of overvoting at the constituency level go down by about 3.15 percentage points. This corresponds to about a 43 percent reduction in a constituency's rate of overvoting. The second two columns separate the constituencies according to whether they are competitive or stronghold. We lose substantial statistical power in these models, but the magnitude of the coefficients indicates that observers lead to reductions in overvoting in all types of constituencies. We find similar patterns with respect to voter turnout. These findings offer evidence that, despite the displacement effects of observers, increasing the number of observers in a constituency reduces overall rates of fraudulent activity.

Table 6 about here

10 Discussion and Conclusions

Ghana is well known as one of sub-Saharan Africa's most stable new democracies. The country is routinely applauded by regional and international bodies for conducting free, fair and relatively peaceful elections. However, our investigation finds evidence of electoral irregularities in Ghana's December 2012 election. Using two indicators of electoral fraud — overvoting and suspiciously high turnout levels — we document irregularities at just over 8 percent of the polling stations in a random sample of slightly over $2,000$ polling stations.

Our results show that domestic election observers substantially reduce the probability of fraudulent activity at the polling centers where observers are stationed. In contrast to the existing literature on observers, we leverage our randomized saturation design to adjust our estimates of observers' causal effects to correct for the bias generated by observers' impact on unobserved stations. We additionally take advantage of the randomization of observer saturation at the constituency level to show that increasing the share of observed polling places in a constituency results in overall reductions of fraudulent activity. Taken together, these findings endorse the conclusion that the deployment of thousands of election observers to polling stations on election day substantially promotes election integrity in Ghana.

That observers reduce overvoting and turnout at the polling stations to which they are deployed is consistent with our interpretation that observers reduce fraud. Yet it could also be the case that observers simply reduce administrative negligence or incompetence. That we find spillover, however, suggests that the irregularities we uncover are the product of deliberate and coordinated political actions across polling stations, and not merely administrative error or incompetence. It may be that observers reduce administrative negligence where they are located, but we would not expect observers to displace negligence to nearby stations or to stations in the same constituency.

We find that the direction of spillover effects varies with the partisan environment in Ghana. In the political party strongholds, observers displace fraudulent activity to unobserved stations in

the same constituency. In more electorally competitive constituencies, we find no evidence of this type of displacement and find evidence of a slight deterrent effect on our two outcomes. It is common knowledge in Ghana that domestic election observers are assigned to a polling station for the entire day, and that a much smaller group of CODEO supervisors visit multiple polling stations. Political parties, whose agents are present in virtually all polling stations across the nation, therefore know early on election day which polling stations are under observation, and which are not. Why is it that parties are able to respond effectively to this information in their strongholds, while in the competitive constituencies they appear unable to do so?

The ability of political parties to respond to observer presence in their strongholds is a likely product of the advantages of greater social penetration and a denser network of willing collaborators — voters, election officials and party agents — in these areas. Our data cannot distinguish whether overvoting or abnormally high turnout figures are the result of political parties shepherding unregistered or double voters to unobserved polling stations or if electoral officials within these stations take advantage of the fact that they were not being monitored to swing results in the favor of their preferred party by manipulating the results. Discussions with CODEO officials suggest that the latter is more likely. As local election officials are generally recruited from the communities in which they are stationed on election day, parties are likely able to use their social penetration to recruit and co-opt electoral officials. In addition, election officials may feel social pressure to use nefarious tactics or turn a blind eye to electoral malfeasance in communities where one of the parties is dominant. News reports on election day confirm that in some cases officials manning polling stations were suspected of favoring their preferred political party.[22]

In competitive constituencies, on the other hand, parties do not enjoy such strong social penetration. Additionally, parties may be better able to police one another in such areas. Political parties, even if they are responsible for election day irregularities, as our results suggest, can be agents who reduce the fraud of their competitors. In the political party stronghold areas, which are

[22]See `http://www.ghanaweb.com/GhanaHomePage/NewsArchive/artikel.php?ID=258814`.

often hostile environments for the agents of the opposition party, such informal policing may be more difficult or even impossible. An implication is that political competition between political parties should reduce coordinated fraud because in such environments political parties have the ability to informally police rivals.

These differential patterns of spillover yield broader implications for research on the conditions in which parties are able to coordinate fraud. The finding that displacement effects are concentrated in uncompetitive political constituencies complements the results of Ziblatt (2009), which reports that land inequality in 19th century Germany predicts electoral fraud because in such settings elites are able to use their social and economic power to capture local electoral administration. Our findings also relate to the argument of Weidmann and Callen (2013), which suggests that the loyalty networks of incumbents in Afghanistan facilitate election fraud through their positions in the electoral administration. That it is widely believed in Ghana that political parties are able to co-opt electoral officials and opposition party agents in their strongholds suggests that a similar dynamic is driving the coordinated fraudulent response to observers that we document. An implication is that the intense partisan competition that characterizes Ghanaian politics may be a necessary, but not sufficient condition for parties to coordinate fraud. They may also require social penetration and power that permits them to capture the local electoral process.

Our analysis of spillover also yields methodological implications. For scholars conducting experiments assessing the impact of interventions designed to improve governance, our results highlight the need to directly assess political actors' strategic responses to these interventions. In illustrating the displacement effects of observers, we show that such strategic responses can influence outcomes in units not designated for the intervention. We additionally show that the direction and magnitude of spillover can be sensitive to the social and political context in which an experiment is conducted. These spillover effects are likely to emerge in other types of governance experiments, including for example those that assess the impact of information campaigns or citizen monitoring efforts. While these types of spillovers from treated to untreated units pose

a challenge to causal estimation, we show in this paper that they also present opportunities for learning about politics. In our case, we were able to make inferences about the conditions in which parties are able to strategically coordinate fraud. Such opportunities may present themselves in other types of experiments as well.

Our study also highlights the importance of studying the net, or equilibrium effects, of experimental interventions. Field experiments in political science and economics have often been criticized for their failure to do so. We show in this paper how randomizing treatment concentration across geographic units — in our case, the saturation of observers in constituencies — can be a useful tool in this endeavor.

Finally, our findings generate an additional question: how do election observers reduce fraud? We can imagine two separate channels. The first is through social sanctioning. Where observers are present, the actors who engage in fraud at polling places — political party agents, electoral commission officials, party sympathizers, and so on — who themselves are often members of the communities in which the polling stations are located, choose to refrain from engaging in fraud because they fear social sanctioning from a community committed to democratic values. In this model, observers facilitate bottom-up enforcement of democratic norms. An alternate channel is legal-judicial. According to this view, observers represent a higher probability that legal officials will be alerted to violations of election law and penalties imposed. Political parties refrain from fraudulent activities where an election observer is present because they do not want such activities to be reported to legal agencies that could potentially impose penalties. These represent two different enforcement mechanisms. Perhaps both are necessary for new democracies to acquire longterm stability.

References

Baird, Sarah, Aislinn Bohren, Craig McIntosh and Berk Ozler. 2012. "Designing Experiments to Measure Spillover and Threshold Effects." Unpublished paper.

Bensel, Richard Franklin. 2004. *The American ballot box in the mid-nineteenth century*. Cambridge University Press.

Boafo-Arthur, Kwame. 2006. The 2004 General Elections: An Overview. In *Voting for Democracy in Ghana: The 2004 Elections in Perspective*, ed. Kwame Boafo-Arthur. Vol. 1 Freedom Publications.

Enikolopov, Ruben, Vasily Korovkin, Maria Petrova, Konstantin Sonin and Alexei Zakharov. 2013. "Field Experiment Estimate of Electoral Fraud in Russian Parliamentary Elections." *Proceedings of the National Academy of Sciences* 110(2):448–452.

Gerber, Alan and Donald Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton.

Herron, Erik S. 2010. "The Effect of Passive Observation Methods on Azerbaijan's 2008 Presidential Election and 2009 Referendum." *Electoral Studies* 29(3):417–424.

Hidalgo, F. Daniel and Simeon Nichter. 2013. "Voter Buying: Shaping the Electorate Through Clientelism." Unpublished paper.

Hyde, Susan D. 2007. "The Observer Effect in International Politics: Evidence From a Natural Experiment." *World Politics* 60(1):37–63.

Hyde, Susan D. 2010. "Experimenting in Democracy: International Observers and the 2004 Pesidential Elections in Indonesia." *Perspectives on Politics* 8(2):511–27.

Hyde, Susan D. 2011. *The Pseudo-Democrat's Dilemma: Why Election Observation Became an Inrnational Norm*. Cornell University Press.

Hyde, Susan D and Angela O'Mahony. 2010. "International scrutiny and pre-electoral fiscal manipulation in developing Countries." *The Journal of Politics* 72(03):690–704.

Ichino, Naomi and Matthias Schündeln. 2012. "Deterring or Displacing Electoral Irregularities? Spillover Effects of Observers in a Randomized Field Experiment in Ghana." *Journal of Politics* 74(1):292–307.

Jockers, Heinz, Dirk Kohnert and Paul Nugent. 2010. "The successful Ghana election of 2008: A convenient myth?" *The Journal of Modern African Studies* 48(1):95–115.

Keefer, Philip. 2002. "DPI2000: Database of Political Institutions: Changes and Variable Definitions." *March. The World Bank* .

Kelley, Judith. 2011. "Data on International Election Monitoring: Three Global Datasets on Election Quality, Election Events and International Election Observation." ICPSR1461-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. doi:10.3886/ICPSR31461.v1.

Kelley, Judith G. 2012. *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails*. Princeton: Princeton University Press.

Lehoucq, Fabrice Edouard. 2003. "Electoral Fraud: Causes, Types, and Consequences." *Annual Review of Political Science* 6(1):233–56.

Molina, Iván and Fabrice Edouard Lehoucq. 1999. "Political competition and electoral fraud: A Latin American case study." *Journal of Interdisciplinary History* 30(2):199–234.

Rubin, Donald B. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of educational Psychology* 66(5):688.

Simpser, Alberto and Daniela Donno. 2012. "Can International Election Monitoring Harm Governance?" *The Journal of Politics* 74(02):501–513.

Sjoberg, Fredrik M. 2012. "Making Voters Count: Evidence from Field Experiments about the Efficacy of Domestic Election Observation." Unpublished paper.

Smith, Daniel A. 2002. "Consolidating Democracy? The Structural Underpinnings of Ghana's 2000 Elections." *The Journal of Modern African Studies* 40(4):621–50.

Weidmann, Nils B. and Michael Callen. 2013. "Violence and Election Fraud: Evidence from Afghanistan." *British Journal of Political Science* 43(01):53–75.

Whitfield, James. 2009. "'Change for a Better Ghana': Party Competition, Institutionalization and Alternation in Ghana's 2008 Elections." *African Affairs* 108(433):621–41.

Ziblatt, Daniel. 2009. "Shaping Democratic Practice and the Causes of Electoral Fraud: The Case of Nineteenth-Century Germany." *American Political Science Review* 103(1):1–21.

Table 1: Descriptive Statistics of Measures of Potential Fraud

	(1) Full Sample	(2) Observed	(3) Unobserved	(4) Swing	(5) NPP Stronghold	(6) NDC Stronghold	(7) Urban	(8) Rural
Overvoting	0.042	0.026	0.070	0.044	0.045	0.031	0.043	0.040
	(0.200)	(0.158)	(0.255)	(0.206)	(0.208)	(0.173)	(0.204)	(0.196)
Turnout Rate	0.834	0.819	0.860	0.825	0.871	0.813	0.818	0.849
	(0.310)	(0.231)	(0.413)	(0.280)	(0.296)	(0.384)	(0.361)	(0.251)
Observations	2026	1273	753	1121	471	434	977	1049

Standard deviations in parantheses.

Table 2: Polling Station Level Covariate Balance

	Mean Control	Mean Observed	Difference	P-Value
NPP Presidential Vote 2008	.357	.355	.002	.864
NDC Presidential Vote 2008	.436	.433	.002	.881
NPP Parliamentary Vote 2008	.383	.358	.025	.087
NDC Parliamentary Vote 2008	.408	.414	-.006	.694
Poverty index	.984	.963	.02	.23
Electricity	1.154	1.129	.025	.286
Medicine	.891	.905	-.014	.514
Sufficient Food	.881	.842	.038	.1
Cash Income	1.008	.976	.032	.126
No Formal Schooling	.147	.15	-.003	.793
Completed Primary Schooling	.685	.708	-.022	.11
Post Primary Schooling	.511	.537	-.026	.088
Formal House	.172	.178	-.006	.626
Concrete Permanent House	.41	.422	-.012	.427
Concrete and Mud House	.224	.215	.008	.504
Mud House	.187	.179	.008	.494

Note: Data are from a post-election survey conducted in the communities around each polling station in the sample (N=6,000). P-values are calculated from two-tailed difference-of-means tests.

Table 3: Observer Effects on Indicators of Electoral Fraud

	Overvoting		Turnout	
	(1)	(2)	(3)	(4)
Observer Present	-0.045***	-0.046***	-0.052***	-0.055***
	(0.011)	(0.011)	(0.014)	(0.016)
Medium Saturation		-0.007		-0.020
		(0.018)		(0.023)
High Saturation		-0.003		0.000
		(0.016)		(0.024)
Competition		0.014		-0.013
		(0.011)		(0.018)
Urban		0.005		-0.017
		(0.011)		(0.019)
Constant	0.071***	0.068***	0.880***	0.903***
	(0.011)	(0.019)	(0.015)	(0.022)
Observations	1,917	1,917	1,917	1,917
R-squared	0.012	0.013	0.007	0.009

Robust standard errors clustered by constiuency in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Adjusted Observer Effect Estimates

	Low Saturation	Linear
Panel A: Overvoting		
Unadjusted Treatment Effect	-.045	-.045
Estimated Pure Control	.07	.063
Average Bias in Control	.001	.008
Adjusted Treatment Effect	-.044	-.037
Panel B: Turnout		
Unadjusted Treatment Effect	-.052	-.052
Estimated Pure Control	.883	.853
Average Bias in Control	-.003	.027
Adjusted Treatment Effect	-.055	-.025

Note: The table presents adjusted estimates of the direct observer effect, correcting for spillover onto control polling stations. In the “low saturation" columns, we use the mean of the control group in the lowest saturation constituencies as an estimate of the pure control outcome. In the “linear" columns, we linearize the relation between observer saturation and each outcome and take the predicted value in the control group where saturation is equal to zero as an estimate of the pure control outcome. The difference between the estimated pure control and the control used in the naive treatment effect estimator gives the estimated average bias in the control group. We use this estimate to adjust our treatment effect estimates to account for the bias associated with spillover.

Table 5: Spatial Spillover Effects of Observers on Indicators of Electoral Fraud

PANEL A: Overvoting	Full Sample	Strongholds	Competitive
Spillover Effect on Unobserved in the same EA	**0.066**	**0.097**	0.049
	(0.008, 0.103)	**(0.029, 0.147)**	(-0.018, 0.094)
Spillover Effect on Observed in the same EA	0	0.031	-0.021
	(-0.043, 0.035)	(-0.043, 0.083)	(-0.072, 0.02)
PANEL B: Turnout	Full Sample	Strongholds	Competitive
Spillover Effect on Unobserved	**0.074**	**0.107**	0.045
	(0.014, 0.132)	**(0.046, 0.196)**	(-0.031, 0.106)
Spillover Effect on Observed	-0.003	0.065	-0.04
	(-0.081, 0.052)	(-0.091, 0.132)	(-0.119, 0.023)

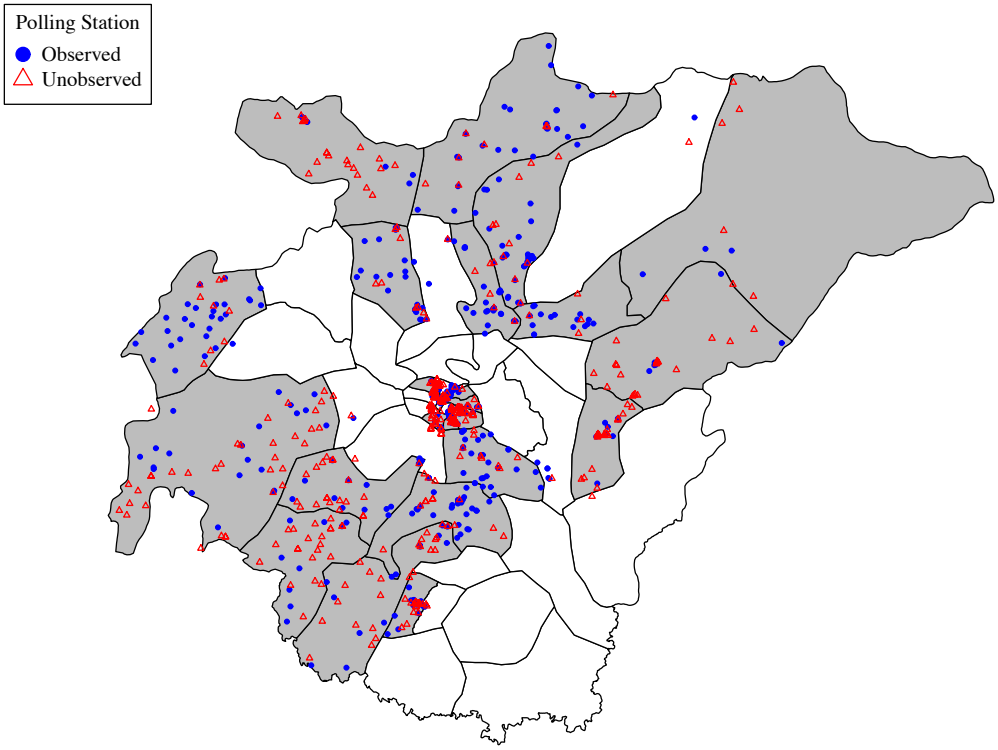
Note: Lower and upper bounds of 95 percent confidence intervals calculated using randomization inference with an assumption of constant effects across all units in parentheses. Estimates in bold are those for which the 95 percent confidence interval does not contain zero.

Table 6: Observer Saturation and Constituency Level Indicators of Fraud

	Overvoting			Turnout		
	Full Sample	Competitive	Stronghold	Full Sample	Competitive	Stronghold
	(1)	(2)	(3)	(4)	(5)	(6)
Observer Saturation	-0.061*	-0.034	-0.086	-0.050	-0.074	-0.023
	(0.033)	(0.045)	(0.050)	(0.051)	(0.064)	(0.089)
Competition				-0.012		
				(0.020)		
Polling Station Density	0.002	0.018	-0.022	-0.012	-0.001	-0.033
	(0.012)	(0.016)	(0.020)	(0.019)	(0.023)	(0.036)
Constant	0.076***	0.052*	0.101***	0.885***	0.884***	0.879***
	(0.020)	(0.028)	(0.031)	(0.031)	(0.039)	(0.054)
Observations	60	38	22	60	38	22
R-squared	0.056	0.052	0.214	0.035	0.037	0.053

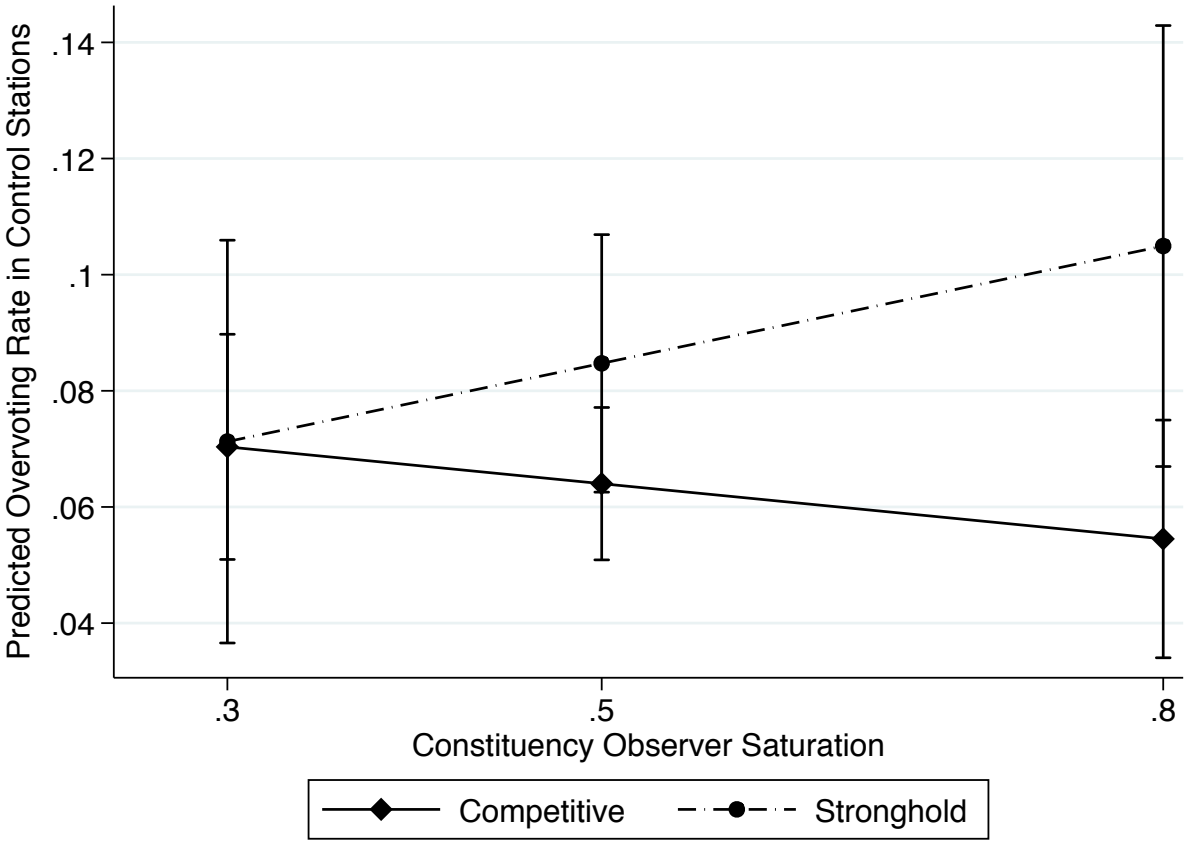
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 1: Sampled Constituencies and Polling Stations in the Ashanti Region



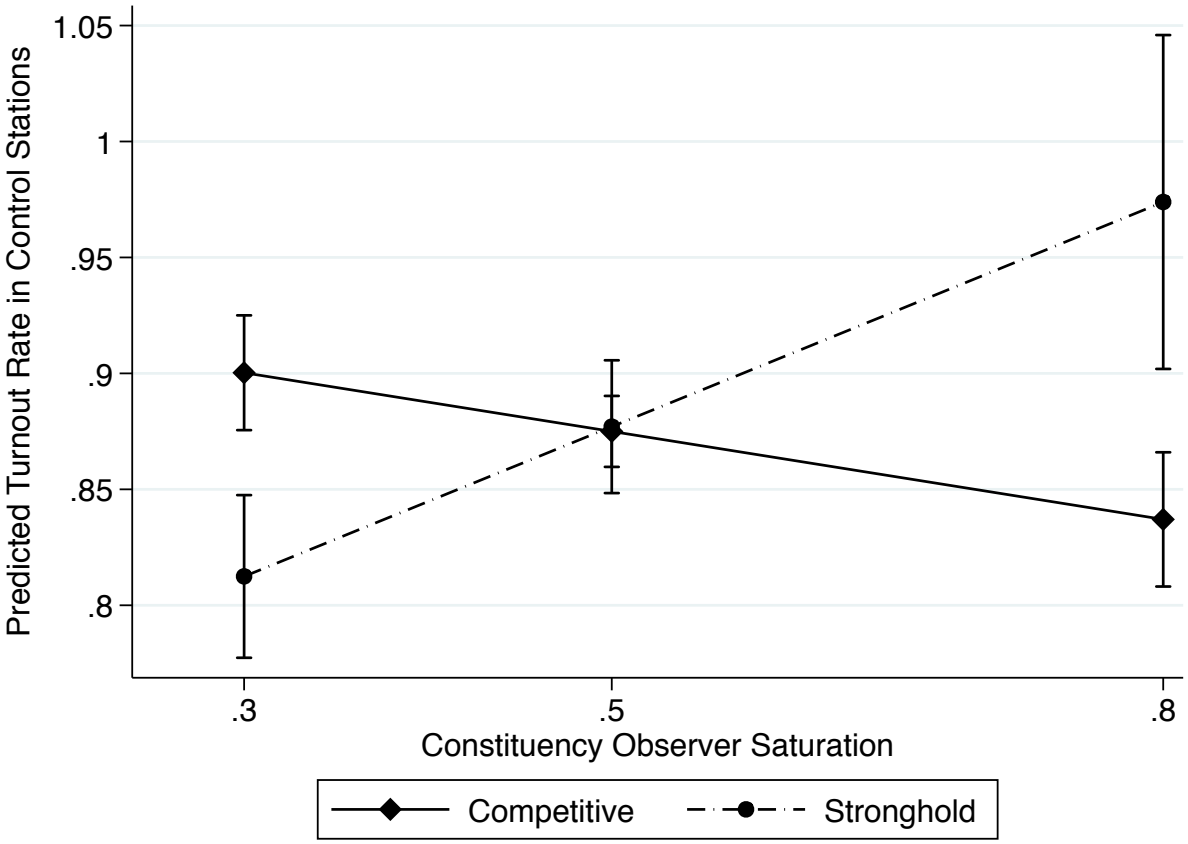
Shaded constituencies are in the sample. A few polling station points fall outside of shaded constituencies in the map because of small changes to constituency boundaries. We use the most up-to-date constituency shapefile we could access, which does not reflect all of these changes.

Figure 2: Overvoting in Unobserved Polling Stations: Spillover in Stronghold versus Competitive Constituencies



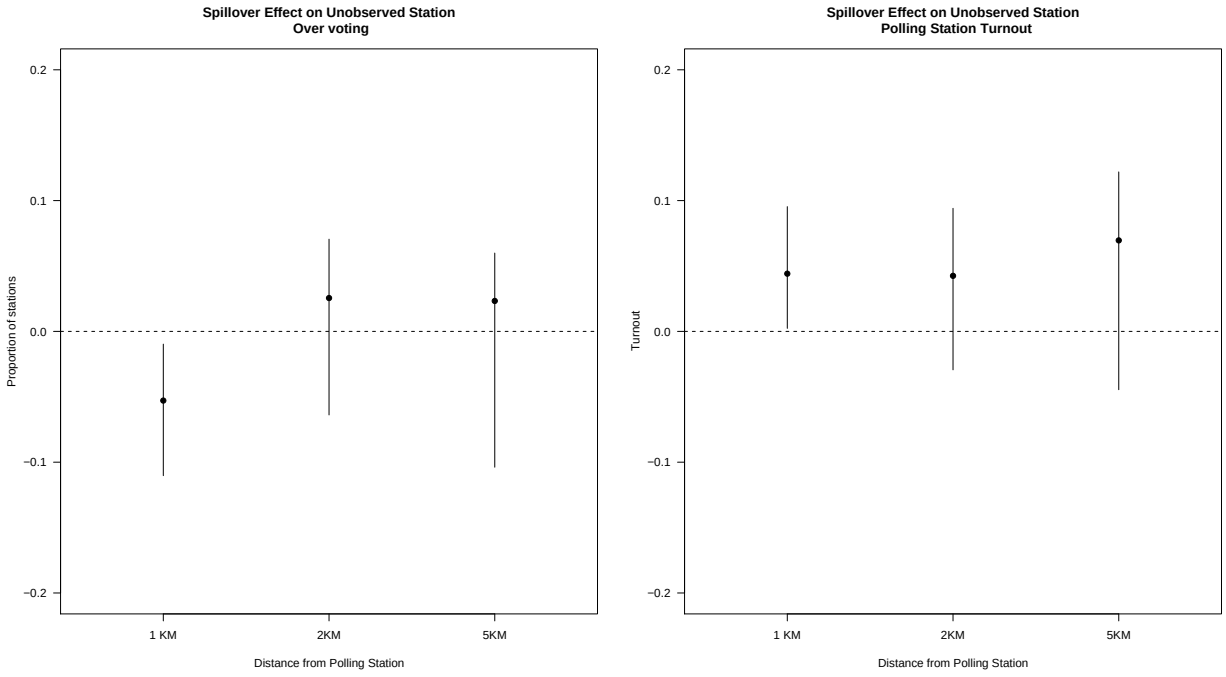
Estimates are generated from the model in column 1 of Table A.1 in the appendix. Plot displays estimated overvoting rates in unobserved (control) polling stations by constituency type and observer saturation. Increases in the overvoting rate as saturation increases is evidence that observers displace overvoting to unobserved polling places within the same constituency. Decreases in the overvoting rate as saturation increases is evidence that observers deter overvoting in unobserved polling places within the same constituency.

Figure 3: Turnout in Unobserved Polling Stations: Spillover in Stronghold versus Competitive Constituencies



Estimates are generated from the model in column 2 of Table A.1 in the appendix. Plot displays estimated turnout rates in unobserved (control) polling stations by constituency type and observer saturation. Increases in the turnout rate as saturation increases is evidence that observers displace turnout fraud to unobserved polling places within the same constituency. Decreases in the turnout rate as saturation increases is evidence that observers deter turnout fraud in unobserved polling places within the same constituency.

Figure 4: Spatial Displacement Effects in Ashanti Region



A Supplementary Materials

A.1 Ballot Stuffing Results

Table A.1: Observer Effects on Ballot Stuffing

	(1) Ballot stuffing	(2) Ballot stuffing
Observer Present (OP)	-0.039	-0.037
	(0.026)	(0.025)
Medium Saturation		0.022
		(0.024)
High Saturation		0.010
		(0.016)
Competition		0.019
		(0.018)
Urban		-0.007
		(0.017)
Constant	0.071***	0.052**
	(0.024)	(0.021)
Observations	2,004	2,004
R-squared	0.008	0.011

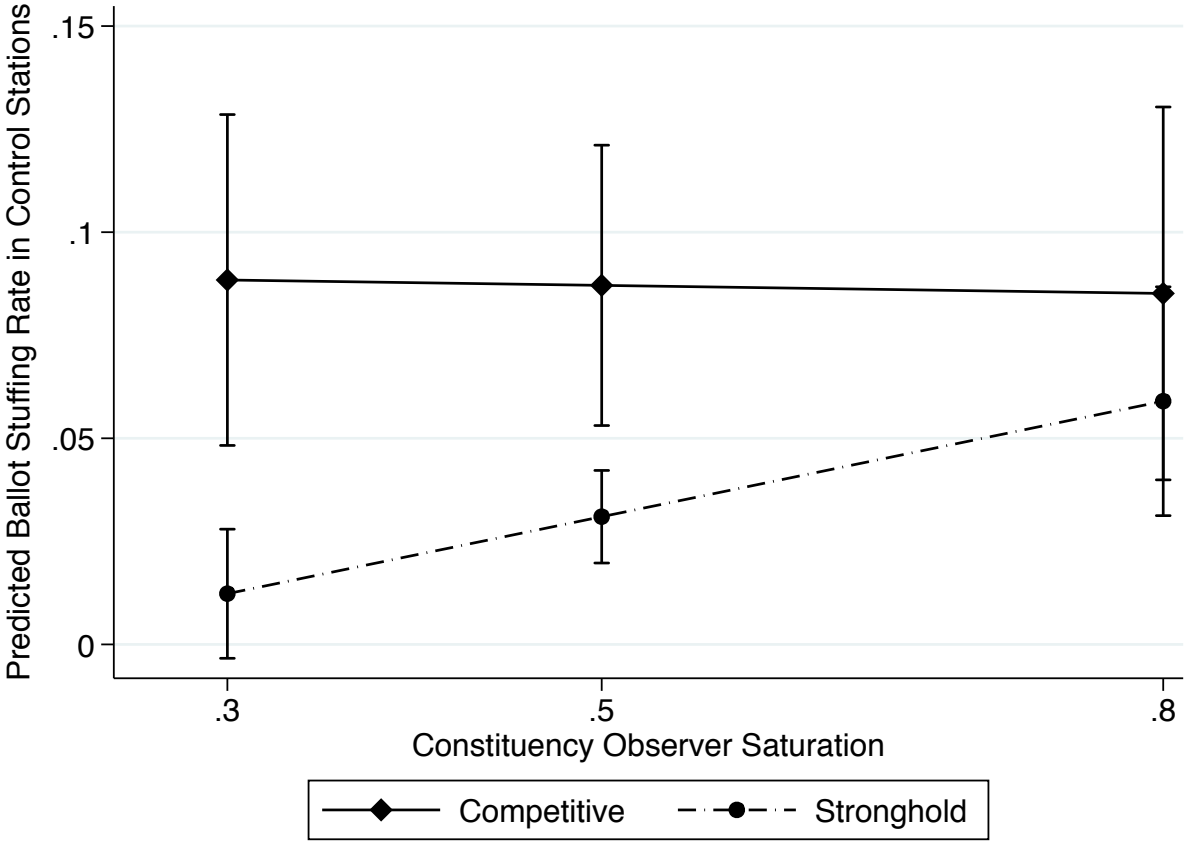
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.2: Adjusted Observer Effect Estimates on Ballot Stuffing

	Low Satuarion	Linear
Naive Treatment Effect	-.039	-.039
Estimated Pure Control	.051	.064
Average Bias In Control	.02	.007
Adjusted Treatment Effect	-.019	-.032

Note: The table presents adjusted estimates of the direct observer effect, correcting for spillover onto control polling stations. In the “low saturation" columns, we use the mean of the control group in the lowest saturation constituencies as an estimate of the pure control outcome. In the “linear" columns, we linearize the relation between observer saturation and each outcome and take the predicted value in the control group where saturation is equal to zero as an estimate of the pure control outcome. The difference between the estimated pure control and the control used in the naive treatment effect estimator gives the estimated average bias in the control group. We use this estimate to adjust our treatment effect estimates to account for the bias associated with spillover.

Figure A.1: Ballot Stuffing in Unobserved Polling Stations: Spillover in Stronghold versus Competitive Constituencies



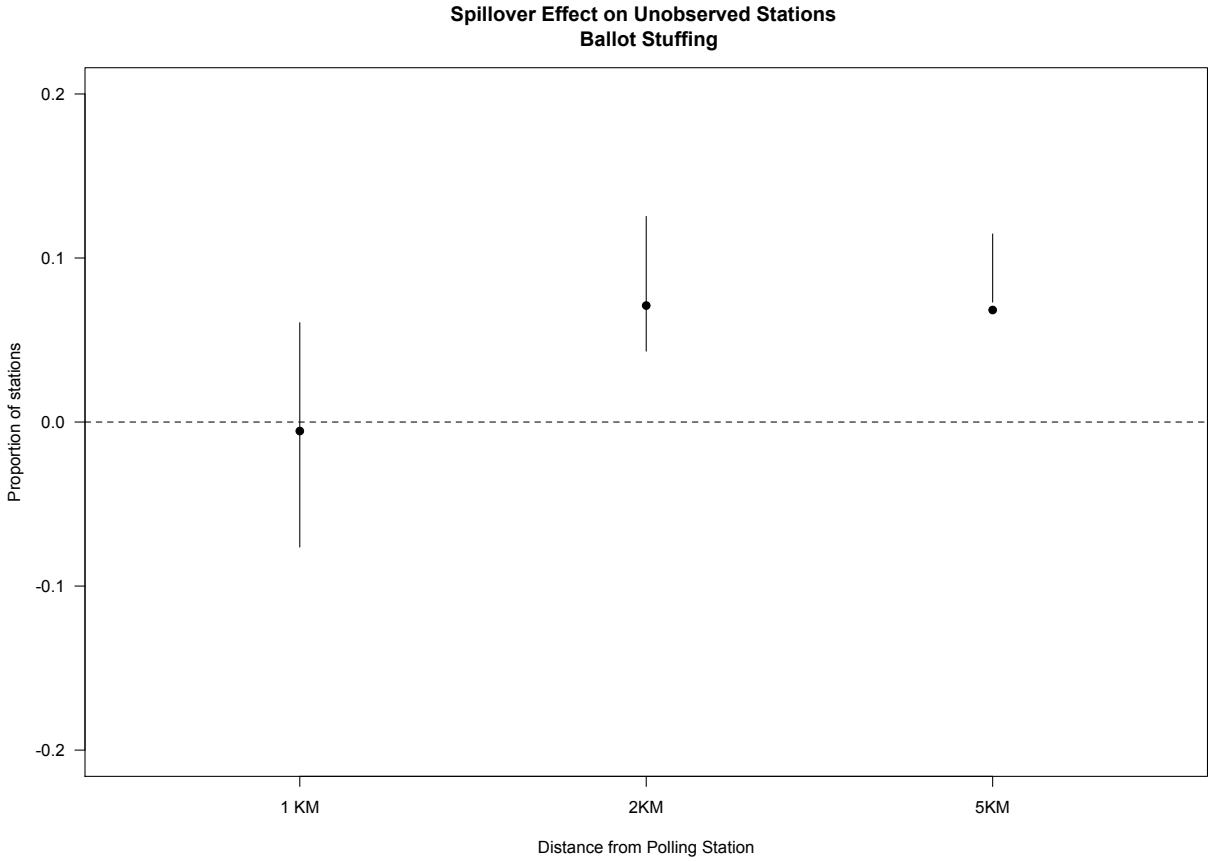
Plot displays estimated ballot stuffing rates in unobserved (control) polling stations by constituency type and observer saturation. Increases in the ballot stuffing rate as saturation increases is evidence that observers displace ballot stuffing fraud to unobserved polling places within the same constituency. Decreases in the ballot stuffing rate as saturation increases is evidence that observers deter ballot stuffing in unobserved polling places within the same constituency.

Table A.3: Spatial Spillover Effects of Observers on Ballot Stuffing

Ballot Stuffing	Full Sample	Strongholds	Competitive
Spillover Effect on Unobserved	0.043	0.03	0.052
	(-0.033, 0.087)	(-0.059, 0.072)	(-0.043, 0.109)
Spillover Effect on Observed	0.025	0.047	0.01
	(-0.026, 0.062)	(-0.038, 0.085)	(-0.05, 0.055)

Note: Lower and upper bounds of 95 percent confidence intervals calculated using randomization inference with an assumption of constant effects across all units in parentheses.

Figure A.2: Spatial Displacement Effects on Ballot Stuffing in Ashanti Region



A.2 Spillover Effects in Strongholds and Competitive Constituencies

Table A.4: Estimates of Spillover Effects in Strongholds and Competitive Constituencies

	(1) Overvoting	(2) Turnout
Observer Present	-0.059*	-0.103**
	(0.034)	(0.041)
Saturation	-0.032	-0.126
	(0.061)	(0.089)
Stronghold Constituency	-0.029	-0.223**
	(0.074)	(0.097)
Observer Present * Stronghold	0.050	0.215**
	(0.066)	(0.102)
Saturation * Observer Present	0.049	0.106
	(0.066)	(0.076)
Saturation * Stronghold	0.099	0.449**
	(0.131)	(0.205)
Observer Present * Saturation * Stronghold	-0.154	-0.414*
	(0.127)	(0.212)
Urban	0.006	-0.016
	(0.011)	(0.018)
Constant	0.077**	0.946***
	(0.035)	(0.049)
Observations	1,917	1,917
R-squared	0.015	0.015

Robust standard errors clustered by constituency in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Utility, Risk, and Demand for Incomplete Insurance: Lab Experiments with Guatemalan Cooperatives

Alain de Janvry
University of California, Berkeley

Craig McIntosh
University of California, San Diego

Felix Povell
Kreditanstalt für Wiederaufbau (KfW), Frankfurt

Elisabeth Sadoulet
University of California, Berkeley

March 2013

JEL Codes: D01, D81, O16

Abstract.
We play a series of incentivized laboratory games with risk-exposed coffee farmers in Guatemala to understand demand for imperfect insurance. We focus on three critical debates in the insurance literature, each of which has important implications for the optimal design of insurance products in the developing world. First, the role of background risk in driving demand; second, the ambiguity between risk aversion and imperfect insurance demand if purchasing insurance can make the worst state of the world worse; and third, the potential for group loss adjustment to improve the performance of index insurance contracts. We find insurance demand to be increasing in background risk, even when the uninsured risk has a negative correlation with the insured risk. Our data strongly confirm the idea that demand for imperfect insurance is sensitive to events in the left tail, and that the possibility of contract non-performance in the worst state of the world strongly dampens demand. Our results on the promise of group insurance are more mixed; we find that individuals in this context are willing to pay less for group insurance than individual (all else held constant), but that they both recognize and are willing to pay for the ability of the group to loss adjust. Overall, our results suggest that demand for partial insurance should be robust in risky parts of the world, that the design of index insurance products should pay particular attention to matching payouts to states with very large losses, and that group insurance can help to overcome some of the problems inherent in index insurance products.

Keywords: behavior under uncertainty, risk aversion, insurance

Thanks to BASIS/USAID for financial support, and to Ethan Ligon for useful comments.