

NBER WORKING PAPER SERIES

PERSONALITIES AND PUBLIC SECTOR PERFORMANCE:  
EVIDENCE FROM A HEALTH EXPERIMENT IN PAKISTAN

Michael Callen  
Saad Gulzar  
Ali Hasanain  
Yasir Khan  
Arman Rezaee

Working Paper 21180  
<http://www.nber.org/papers/w21180>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
May 2015

We thank Farasat Iqbal (Punjab Health Sector Reforms Project) for championing and implementing the project and, Asim Fayaz, and Zubair Bhatti (World Bank) for designing the smartphone monitoring program. Support is provided by the International Growth Centre (IGC) state capabilities program, the IGC Pakistan country office, and the University of California Office of the President Lab Fees Research Program Grant #235855. Callen was supported by grant #FA9550-09-1-0314 from the Air Force Office of Scientific Research. We thank Tahir Andrabi, Eli Berman, Ali Cheema, Julie Cullen, Clark Gibson, Naved Hamid, Gordon Hanson, Asim Ijaz Khwaja, Jennifer Lerner, Jane Mansbridge, Edward Miguel, Craig McIntosh, Ijaz Nabi, Rohini Pande, Christopher Woodruff and seminar participants at UC Berkeley, UC Los Angeles Anderson, UC San Diego, Paris School of Economics, New York University, University of Washington, Harvard Kennedy School, and participants at the IGC Political Economy Group, Development and Conflict Research (DACOR), Pacific Development (PacDev), New England Universities Development Consortium (NEUDC), Southern California Conference in Applied Microeconomics (SoCCAM), Bay Area Behavioral and Experimental Economics Workshop (BABEEW), and Symposium on Economic Experiments in Developing Countries (SEEDEEC), and the Bureau for Research and Economic Analysis of Development (BREAD) conferences for insightful comments. Excellent research assistance was provided by Muhammad Zia Mehmood. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Michael Callen, Saad Gulzar, Ali Hasanain, Yasir Khan, and Arman Rezaee. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Personalities and Public Sector Performance: Evidence from a Health Experiment in Pakistan  
Michael Callen, Saad Gulzar, Ali Hasanain, Yasir Khan, and Arman Rezaee  
NBER Working Paper No. 21180  
May 2015  
JEL No. C93,D02,D73,H11,O31

**ABSTRACT**

This paper provides evidence that the personality traits of policy actors matter for policy outcomes in the context of two large-scale experiments in Punjab, Pakistan. Three results support the relevance of personalities for policy outcomes. First, doctors with higher Big Five and Perry Public Sector Motivation scores attend work more and falsify inspection reports less. Second, health inspectors who score higher on these personality measures exhibit a larger treatment response to increased monitoring. Last, senior health officials with higher Big Five scores are more likely to respond to a report of an underperforming facility by compelling better subsequent staff attendance.

Michael Callen  
Harvard Kennedy School  
79 JFK Street  
Cambridge, MA 02138  
and NBER  
michael\_callen@hks.harvard.edu

Saad Gulzar  
New York University  
Department of Politics  
19 West 4th St. 2nd Floor  
New York, NY 10012  
saad.gulzar@nyu.edu

Ali Hasanain  
University College, Oxford  
And  
Lahore University of Management Sciences  
Sector U, DHA  
Lahore 54792, Pakistan  
ali.hasanain@univ.ox.ac.uk

Yasir Khan  
Lahore University of Management Sciences  
Sector U, DHA  
Lahore 54792, Pakistan  
yasir.khan@theigc.org

Arman Rezaee  
Department of Economics, 0508  
University of California, San Diego  
9500 Gilman Drive  
La Jolla, CA 92093 - 0508  
arezaee@ucsd.edu

# 1 Introduction

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

We consider three questions using a large-scale field experiment designed to improve health worker performance in Punjab, Pakistan.<sup>4</sup> First, do personality measures predict performance under status quo incentives, which are weak? Second, do these measures pre-

---

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

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

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

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

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

In answer to our first question, the Big Five and Perry Public Service Motivation (PSM) measures systematically predict doctor and, to a lesser extent, inspector performance.<sup>5</sup> Doctors who score one standard deviation higher on the measured Big Five trait of conscientiousness, for example, are 5.5 percentage points more likely to be present at work during an unannounced visit. Similarly, health inspectors that score one standard deviation higher on the measured PSM trait of commitment to policymaking are five percentage points less likely to be found colluding with doctors to falsify inspection reports. In addition, health inspectors that score one standard deviation better on a proxy measure of the tendency to procrastinate are 5.8 percentage points more likely to complete each of their assigned inspections in a month.<sup>6</sup> Overall, we find significant positive correlations for four of eleven measured doctor personality traits and doctor attendance, and six of the remaining seven coefficients are also positive.<sup>7</sup> To help understand these coefficients, our back-of-the-envelope estimates indicate

---

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

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

<sup>7</sup>Throughout the paper, we will scale our personality measures such that higher values are normatively better from the perspective of worker performance. We will report results both on individual traits and on summary indices of the Big Five and Public Sector Motivation traits. We acknowledge that focusing on indices rather than specific measures runs counter to the fundamental intention of these measures—to

that replacing the bottom 25 percent of doctors in terms of the aggregate Big Five index with average doctors would result in about 4,650 or four percent more patients being seen by a doctor every month in Punjab’s rural public health facilities.<sup>8</sup> A similar, though weaker, pattern holds with health inspectors, which we discuss in detail below.

To provide evidence on the second question, we designed and implemented a smartphone technology that verifies whether officials are performing regular facility inspections across Punjab, which we evaluated using a randomized control trial spanning the province.<sup>9</sup> We find that a one standard deviation increase in our measured aggregate Big Five index for a government inspector is associated with a 35 percentage point differential increase in inspections in response to treatment.<sup>10</sup>

On the final question, a one standard deviation increase above the mean in our measured aggregate Big Five index of a senior health official is associated with an additional 40 percentage point reduction in doctor absence at a facility managed by the official when the facility’s performance is experimentally flagged for the official’s attention.<sup>11</sup> These offi-

---

identify distinctive features of individual personality. We use indices for two reasons. First, we find that better traits predict better performance irrespective of the specific trait so that indices are useful for brevity. Second, from an econometric perspective, using an index provides a means of dealing with multiple hypothesis testing.

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

<sup>9</sup>Considering the distribution of personality types of agents most affected by an intervention can also help us understand what treatment effects might look like in other settings. On an intuitive level, if a bureaucracy is staffed with workers whose personalities are well-suited to the job, increasing incentives to perform may make very little difference. Conversely, if workers are highly incompatible with their jobs, reforms may induce little additional effort. In line with this intuition, we find suggestive evidence that treatment effects from the monitoring technology are localized to the middle of the personality distribution.

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

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

cials oversee health systems responsible for several million citizens. The degree of our result suggests that improvements at this level of the bureaucracy might be particularly impactful.

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

---

prior to the collection of any data on personalities. Finally, we also find similarly strong results even after correcting our standard errors using the method of Benjamini and Hochberg (1995).

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

<sup>13</sup>These two traits are also the most consistently predictive of performance in our data.

Salgado, 1997). Hogan and Holland (2003) find in a meta-analysis that all five Big Five measures positively predict performance on specific job criteria, and that the predictions become stronger as the job criteria become more specific.<sup>14</sup> Regarding whether traits are fixed, Kautz et al. (2014), in a comprehensive review of the literature, argue that the evidence so consistently supports malleability that non-cognitive attributes should be called “skills”, rather than “traits”, partly to re-orient policy toward the value of investing in these dimensions of human capital.<sup>15</sup>

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

---

<sup>14</sup>There is also more general evidence that the traits of senior executives are important in determining the performance of the entities that they manage. At the firm level, Johnson et al. (1985) find that shareholder wealth is positively correlated with measures of a firm’s executive’s ‘talents’ and ‘decision-making responsibility.’ Bertrand and Schoar (2003) find that a significant extent of the heterogeneity in investment, financial, and organizational practices of firms can be explained by the presence of manager fixed effects. Malmendier et al. (2011) find that overconfidence affects management decisions. At the cross-national level, Jones and Olken (2005) find, using deaths of leaders as exogenous variation, that leaders matter for a country’s growth, especially when constraints on the executive are weak.

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

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

outcomes.

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

The paper proceeds as follows: Section 2 outlines a simple model based on Almlund et al. (2011) to explain how personality traits can affect job task selection and performance. Section 3 provides the institutional details necessary to understand our results. Section 4 outlines our research design and reports results. Section 5 concludes.

## 2 Framework

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

Let our personality measures capture a worker's type,  $\theta$ , with cumulative distribution  $F(\theta)$ . Let performance be the binary decision that a doctor or health inspector makes of whether to attend work. If a worker attends, he receives a fixed salary of  $W$  and incurs a cost of effort  $\lambda(\theta)$ . If a worker shirks, he exerts no effort and receives the fixed salary with probability  $1 - p$  and an arbitrarily small punishment  $c$  with probability  $p$ , as well as an

---

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



outside option of  $Q$ .<sup>18</sup>

## 2.1 Personality Type and Performance

The marginal worker indifferent between working and shirking will satisfy

$$W - \lambda(\theta) = (1 - p)W - pc + Q. \quad (1)$$

If work is less costly for better types ( $\frac{\partial \lambda}{\partial \theta} < 0$ ), then all workers with  $\theta$  greater than that of the marginal worker will choose to work. Equation 1 therefore gives that workers with better personality types are weakly more likely to attend work. This accords with Almlund et al. (2011), in which the authors define traits as features which allow individuals to produce more with a fixed amount of effort.<sup>19</sup>

## 2.2 Personality Type and Responses to Changes to Incentives

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

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

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

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

---

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

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

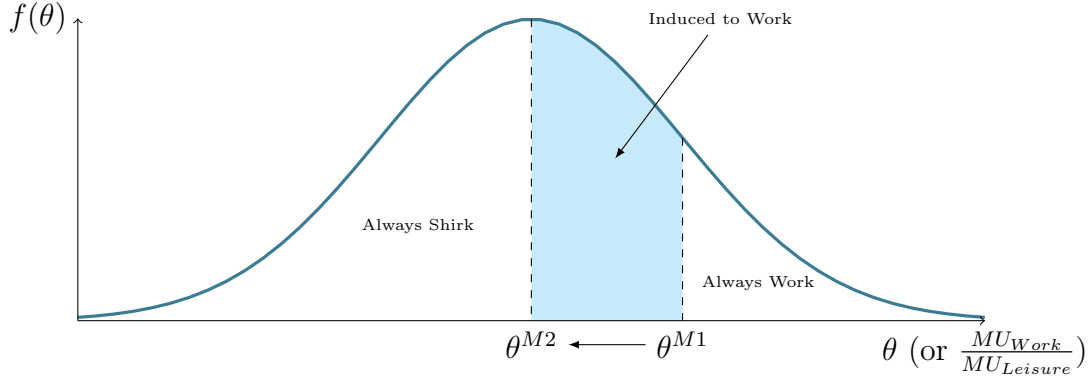


Figure 1: Effect of an Increase in Detection Probability on the Decision to Work or Shirk

We can see this in a simple picture in Figure 1. Let  $\theta^{M1}$  be the marginal worker before an increase in  $p$  and  $\theta^{M2}$  the lower-type marginal worker afterwards. All workers with  $\theta > \theta^{M1}$  continue to work and workers with types in the shaded area  $\theta^{M1} > \theta > \theta^{M2}$  are induced to work by the increase in detection probability. The types induced to work are the highest (best) among those that shirk prior to the shift in  $p$ . Equation 1 therefore also describes how a personality type relates to a reform in incentives.<sup>20</sup>

Here we assume personality traits only affect the cost of effort in an otherwise simple indifference equation. It follows that better personality types are more likely to work ex-ante and that the better types among ex-ante shirkers will be more likely to respond to an increase in incentives. The decision to work is potentially much more complex. For example, personality traits that are useful in the public sector may also increase productivity in the outside option (i.e.,  $Q$  may also be a function of  $\theta$ ). More generally,  $\theta$  might not only capture the single-dimensional productivity gains to personality traits. It may also capture heterogeneity in workers' outside options, in workers' cognitive ability, in workers'

<sup>20</sup>Note that Figure 1 allows us to make two additional points. The first is that the results in this paper, as with all results from randomized interventions, are Local Average Treatment Effects. That is, our intervention may induce some workers to work, but there are some workers that will always work and some that will never work regardless of the intervention. The second point is that the initial position of  $\theta^{M1}$  matters significantly to the size of the impact of an increase in detection probability. This also highlights the importance of the shape of the distribution of personality types, as a very narrow distribution might see different effects than a uniform distribution from an increase in  $p$ . Both the initial position of  $\theta^{M1}$  and the distribution of personality types can be estimated ex-ante, allowing for better targeted policies.

ability to mitigate political pressure from outside their office, and so on. We could deal with this in two ways. Most simply, we could redefine  $\lambda(\theta)$  to include the all of these personality trait-dependent costs and benefits. Then the simple model would encapsulate a richer understanding of these costs and benefits of personality traits, but it would be unable to differentiate these costs and benefits. Second, we could enrich the model by, for example, modeling  $Q$  as a function of  $\theta$ . Without additional and somewhat implausible assumptions, doing so immediately expands the set of predictions to the point where the model is no longer falsifiable. We demonstrate this in Appendix Section A.3.

### 3 Public Health Services in Punjab

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

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

The main performance outcomes in this paper are measured at primary front-line public health clinics, called Basic Health Units (BHUs).<sup>21</sup> BHUs are designed to be the first stop for rural patients seeking medical treatment in government facilities, providing mainly primary services, including out-patient services, neo-natal and reproductive health care, and vaccinations against diseases. Hereafter in this paper, we use the word ‘clinic’ interchangeably to describe BHUs. There are 2,496 BHUs in Punjab.<sup>22</sup> Almost all BHUs are located in rural

---

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

<sup>22</sup>Each Basic Health Unit serves approximately one Union Council (Union Councils are smallest adminis-

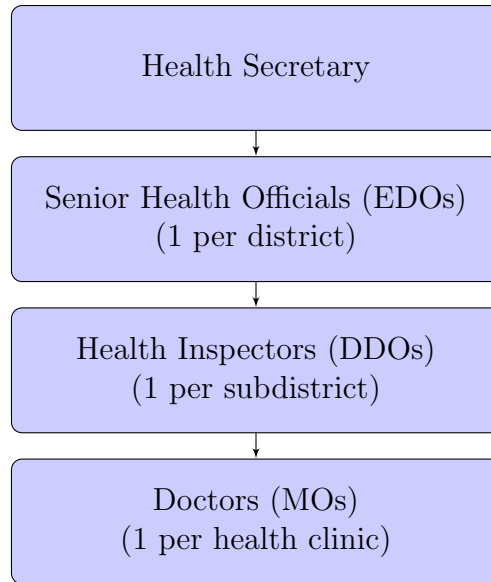


Figure 2: Health Sector Administration in Punjab

and peri-urban areas. Each facility is headed by a doctor, known as the Medical Officer, who is supported by a Dispenser, a Lady Health Visitor, a School Health and Nutrition Supervisor, a Health/Medical Technician, a Mid-wife and other ancillary staff. Officially, clinics are open, and all staff are supposed to be present, from 8AM to 2PM and patients seen in these clinics are required to pay a nominal fee of around \$0.01 USD per visit.

### 3.1 Health Sector Administration

Figure 2 depicts a simplified version of the health administration hierarchy in Punjab. District governments are responsible for managing local health facilities. Each District Department of Health is headed by an Executive District Officer (EDO) who reports both to the official in charge of the district and to two provincial health officials.<sup>23</sup> EDOs are directly supported by several Deputy District Officers (DDOs). DDOs primarily inspect and manage

---

trative units in Pakistan).

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

health facilities.<sup>24</sup> DDOs are required to inspect every clinic in their jurisdiction at least once a month and record information collected during the visit on a standard form. DDOs have the authority to punish a clinic’s absent staff by issuing a formal reprimand, suspending staff, and/or withholding pay (in the case of contract staff). Each Medical Officer is similarly responsible for their own clinic, with proportional duties. Throughout the paper, we will refer to Executive District Officers as senior health officials, Deputy District Officers as inspectors, and Medical Officers as doctors, focusing on their role rather than their title.

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

## 4 Results

In this section, we present three sets of results, each corresponding to one of the three questions laid out in Section 1. First, we study correlations between the measured personality traits of doctors and health inspectors, their job performance (attendance and inspections respectively), and their propensity to collude with one another. Second, we use these measures to predict health inspectors’ response to an experimental intervention which increases the probability of detecting shirking. Finally, we examine whether traits identify which senior health officials react to information about the absence of their subordinates. This analysis relies on a second policy intervention which manipulates the information provided to senior officials about the absence of their subordinates.

---

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

<sup>25</sup>Doctors were not posted at 35 percent of clinics, which means unconditional doctor presence was only 32 percent.

## 4.1 Predicting Doctor and Health Inspector Performance

We first examine if personality measures predict bureaucratic performance under status quo incentives, for doctors and then for health inspectors. We measured personality for 389 doctors in Punjab posted to a representative sample of 850 of the 2,496 rural health clinics in Punjab.<sup>26</sup> We also measured personality for the universe of health inspectors and senior health officials in Punjab, for a total of 101 inspectors and 33 senior health officials. For all 850 clinics in our sample, we also measured attendance during unannounced visits in November 2011, June 2012, and October 2012. Before presenting results, we describe the monitoring intervention that allowed us to collect our data as well as our personality, procrastination, and performance measures.<sup>27</sup>

### 4.1.1 Measuring Personality

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

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

---

<sup>26</sup>306 facilities in this sample have no doctor posted. We omit these clinics from our doctor analyses.

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

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

<sup>29</sup>John et al. (2008) provide the mapping between questions and traits.

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

The Perry Public Service Motivation (PSM) battery is designed to measure intrinsic motivation for public service. Also developed in the 1980s, it comprises a total of 40 questions measuring six traits—attraction to policymaking, commitment to policymaking, social justice, civic duty, compassion, and self-sacrifice.<sup>31</sup> We reproduce both the Big Five and PSM batteries we used in the appendix.<sup>32</sup>

Table 2 reports summary statistics for these measures separately for doctors and health inspectors in treatment and control districts in our randomized control evaluation of a new monitoring technology.<sup>33</sup> There is substantial variation in personality traits across individuals consistent with the original intention of the battery: to capture substantial and important differences in personality types.<sup>34</sup> It is this heterogeneity that allows for the possibility of linking differences in personality to variation in performance.

#### 4.1.2 Measuring Doctor Performance

To obtain measures of performance, we collected primary data on a representative sample of 850 of the 2,496 clinics or Basic Health Units in Punjab. Clinics were selected randomly

---

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

<sup>31</sup>Perry and Wise (1990) and Perry (1996) introduce the battery and Petrovsky (2009) provides a summary of studies using this measure.

<sup>32</sup>Though the survey included is for doctors (Medical Officers), we used the same instrument for health inspectors and senior health officials. We include both the Urdu version that was fielded, as well as a translation of the instrument to English for reference.

<sup>33</sup>We describe the experiment in Subsection 4.1.4 below. We capture these measures after treatment is administered. However, balance on these measures increases our confidence that they are stable over the horizon of the study and unaffected by treatment. The full distributions for these measures are reported in Table A.1. Summary statistics for senior health officials are reported in Table A.2.

<sup>34</sup>Borghans et al. (2008) explain the development of the Big Five.



Figure 3: Locations of Clinics (Basic Health Units) in the Experimental Sample

using an Equal Probability of Selection design, stratified on district and distance between the district headquarters and the clinic. Our estimates of absence are, therefore, self-weighting and require no sampling correction. All districts in Punjab except Khanewal—the technology pilot district—are represented in our data. Figure 3 provides a map of clinics in our experimental sample along with the district boundaries in Punjab.

Information on staff absence, health inspections, and facility usage was collected through three independent and unannounced inspections of these clinics. We visited each facility three times: November 2011, June 2012, and October 2012. Our survey team interviewed and physically verified the presence of the Medical Officer, or doctor, and verified the last health inspection that occurred through written records stored at the facility.<sup>35</sup>

<sup>35</sup>In addition, the attendance of Dispensers, Health/Medical Technicians, Lady Health Visitors, Midwives, and School Health and Nutrition Specialists were also recorded. Survey teams were trained at regional hubs (four in total) by senior enumerator trainers and our team members. Following these trainings, the teams made visits to clinics in their assigned districts and remained in regular contact with their team leaders and our research team. Surveys took three weeks to field for each wave. The attendance sheet for the staff was filled out at the end of the interviews and in private. Inspectors record visits by signing paper registers maintained at the health facility. We measure whether an inspection occurred by interviewing



We have two measures of doctor job performance: (i) whether doctors were present during our unannounced visits, and (ii) a proxy measure of collusion between doctors and health inspectors to falsify reports. We define collusion as a dummy variable coded as one when the doctor is absent during both of our post-treatment unannounced visits and is marked present during every single health inspection during the treatment period.<sup>36</sup> Baseline performance measures for doctors are reported in Table A.1.

### 4.1.3 Personality and Doctor Performance

In Figure 4, Panel A shows that doctors that score one standard deviation above the mean on the Big Five measure of conscientiousness are about five percentage points more likely to be present at work during an unannounced visit. Similarly, two measures of PSM, civic duty and self-sacrifice, are also significantly predictive. Finally, all but one coefficient are positively correlated with doctor attendance. In Panel B, we find that doctor personality measures are even stronger predictors of collusion between health inspectors and doctors. Doctors who score one standard deviation higher on measured civic duty, for example, are about 6 percentage points more likely to be identified as potentially colluding. Ten out of eleven Big Five and PSM traits are highly predictive of collusion, with negative signs.<sup>37</sup>

We draw two lessons from this exercise. First, in Appendix Table A.5, we find that personality is a stronger predictor for doctors than three other plausibly important observables—doctor tenure in the department of health, doctor tenure at the specific health clinic at which

---

facility staff and verifying the register record. Data collection and entry followed back-checks and other validation processes consistent with academic best practice.

<sup>36</sup>The median number of health inspections for each facility in our treatment sample is 12, with a max of 50. The collusion we have in mind occurs when a health inspector calls a doctor before an inspection to alert him to be in attendance. Then, after the health inspector records his presence, the doctor is under very little pressure to attend until he gets another similar phone call from the inspector. Of course, such patterns in the data could arise by chance, though the chance decreases with the number of inspections. As such, we have run all of our collusion analysis using weighted least squares and we find results very similar to those OLS results presented below. Results provided upon request. The strong correlation we find between these measures and personality types also suggests that the proxy is successfully capturing malfeasance. An immediate problem with this proxy is that it partly reflects attendance. We deal with this by adjusting our inference to reflect multiple hypothesis tests using the False Discovery Rate procedure described in Benjamini and Hochberg (1995).

<sup>37</sup>See Appendix Tables A.3 and A.4 for point-estimates.

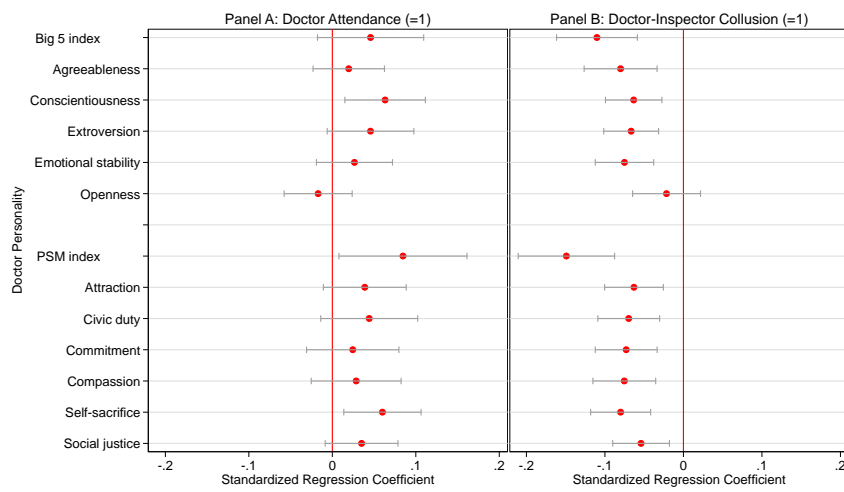


Figure 4: Personality and Performance: Doctors

*Notes:* Each regression coefficient reported comes from a separate regression of the performance measure, Doctor Attendance in Panel A and Doctor-Inspector Collusion in Panel B, on the indicated doctor personality measure. Error bars represent 95 percent confidence intervals, with standard errors clustered at the clinic level. All regressions include Tehsil (county) and survey wave fixed effects. In all cases, personality measures are normalized to have mean zero and standard deviation of one in the sample, and thus the regression coefficients reported can be interpreted as the impact of a one standard deviation increase in a given personality trait or aggregate measure. The sample is restricted to control district clinics for which doctor personality data are available and a doctor is posted. Regressions corresponding to the figure are reported in Appendix Tables A.3 and A.4.

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

Second, the degree of the estimated coefficients is meaningful. As one example, we look at the relationship between doctor attendance and out-patient services. To do so, we first establish a strong positive correlation between doctor presence at their clinic during one of our unannounced visits and reported out-patients seen at that clinic in Appendix Table A.6.

Given this correlation and those in Figure 4, in Appendix Section A.4 we perform a simple back-of-the-envelope exercise to quantify the effect we would expect on out-patients seen if we were to replace the bottom 25 percent of doctors with average doctors in terms of the Big Five index and one specific trait—conscientiousness. We find an expected increase of 4,646 and 9,469 out-patients seen per month from replacing the bottom 25 percent of doctors with average doctors according to the Big Five index and the conscientiousness measure respectively.<sup>38</sup>

#### 4.1.4 Monitoring Intervention

We collected personality data during a larger experimental policy reform that considered audits by government monitors as a solution to the problem of bureaucratic absence. The “Monitoring the Monitors” program replaced the traditional paper-based monitoring system for clinic utilization, resource availability, and worker absence with an android-based smartphone application. In the new system, data generated by health inspections are transmitted to a central database using General Packet Radio Service (GPRS). Data are then aggregated and summary statistics, charts, and graphs are presented in a format designed in collaboration with senior health officials to effectively communicate information on health facility performance. These data are also: (i) geo-tagged, time-stamped, and complemented with facility staff photos to check for reliability; and (ii) available in real time to district and provincial officials through an online dashboard. The objective of this monitoring system is to make the activities of health inspectors available to their senior officials in real time. Figure 5 shows one view of the online dashboard.<sup>39</sup>

---

<sup>38</sup>We treat this exercise speculatively for several reasons. Importantly, it relies on an assumption that the correlation we observe reflects a causal linkage between personality traits and performance. However, in combination with the substantial literature suggesting that personality matters for performance and the related literature on teacher value added (Chetty et al. (2014a); Chetty et al. (2014b); Hanushek (2011); Hanushek and Rivkin (2012); Staiger and Rockoff (2010)), it does not seem too outlandish to argue that performance would improve substantially if the distribution of types were improved.

<sup>39</sup>Application development started in August 2011. After developing the application and linking it to a beta version of the online dashboard, the system was piloted in the district of Khanewal. We remove Khanewal district from the experimental sample. Health administration staff were provided with smartphones and trained to use the application.

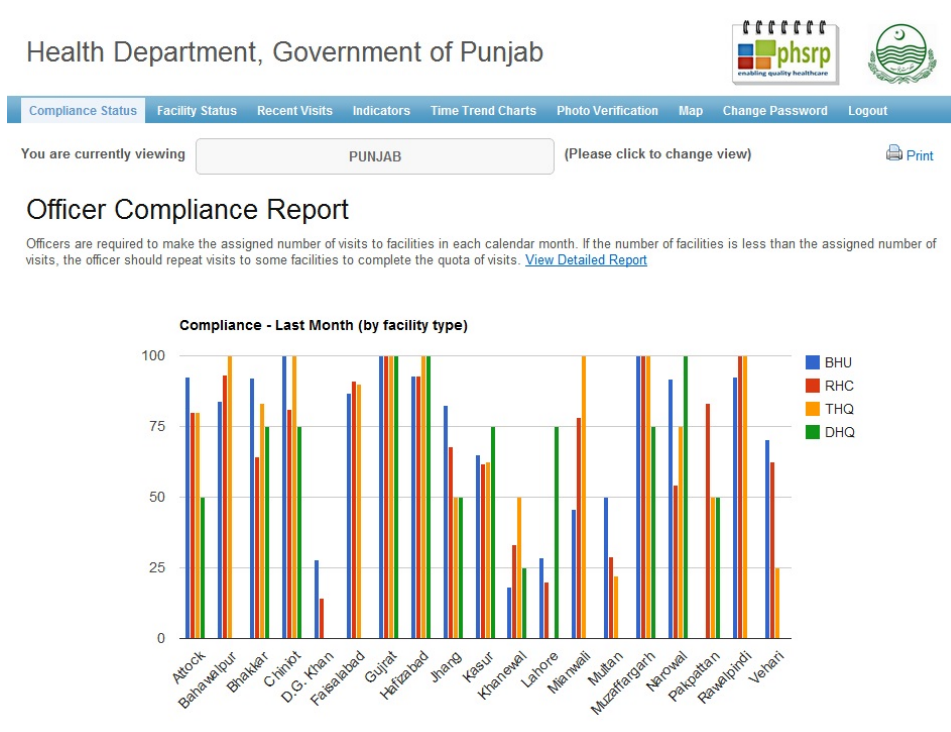


Figure 5: Online Dashboard - Summary of Inspection Compliance by District

In the context of our model above, providing data to senior officials creates a discrete increase to  $p$ , the probability that a health inspector will be caught if he is failing to do his inspections. Prior to Monitoring the Monitors, and in control districts, the paper-based monitoring system severely limits a senior officials ability to monitor inspectors. In treatment districts, on the other hand, reports are immediately and automatically sent up the chain of command, and the required geo-tags, time stamps, and photos serve as instant verification that the inspector and all reported staff are present at the clinic being inspected.<sup>40</sup>

#### 4.1.5 Measuring the Tendency to Procrastinate

A nascent literature uses intertemporal consumption and effort profiles to measure time preference and time inconsistency.<sup>41</sup> Inspectors in Punjab are required to inspect every

<sup>40</sup>See Callen et al. (2013) for the core results from the broad Monitoring the Monitors experiment.

<sup>41</sup>Augenblick et al. (Forthcoming) elicit time preferences based on the intertemporal allocation of non-monetary tasks in the lab. Shapiro (2005) and Kuhn (2013) provide evidence that the intra-month consumption profile of food stamp recipients reflects dynamically-inconsistent planning and better fits a quasi-

facility in their jurisdiction once a month. The intertemporal inspection allocations captured by our smartphone monitoring system reveal patterns indicating a tendency to procrastinate for a majority of our inspectors.

Panel A of Figure 6 depicts the average number of inspections on each day of the month. On the first day of the month, inspectors perform an average of about 0.31 inspections. After the first ten days of the month, average inspections on a given day are roughly 0.8. The time profile of inspections over a month has a positive slope. Several months of data allow estimation of the slope of the intertemporal profile of inspections, providing a proxy measure of each inspector’s tendency to procrastinate. We estimate

$$Inspections_{d,m} = \alpha + \eta \text{ Day of Month}_{d,m} + \delta_m + \epsilon_{d,m} \quad (4)$$

where  $inspections_{d,m}$  is the number of inspections on a given day  $d$  in a month  $m$ ,  $\delta_m$  are fixed effects for each month, and  $Day of Month$  runs from one to 28 depending on the calendar day of the month.<sup>42</sup> Inspectors with a positive  $\eta$  estimate do fewer inspections at the beginning of the month and more at the end as they approach the deadline for their quota, suggesting a tendency to procrastinate.

Panel B of Figure 6 provides a histogram of the estimates of  $\eta$  for 36 inspectors. 29 of these 36 inspectors have positive slope coefficients. The average slope coefficient is 0.014, which indicates that over the course of the month the number of inspections per day increases by about 0.4.

---

hyperbolic model than a standard exponential discounting model.

<sup>42</sup>The effective deadline for inspections is the 28th of the month as senior officials and inspectors meet during the final days of the month to review the month’s inspections. We only include months for which we have complete information for a health inspector and drop holidays. We retain data for 36 health inspectors and have an average of 8.75 months of inspection-level data per inspector. The median number of inspections in a month is 25 and inspectors are responsible for between 4 and 46 facilities with a median of fifteen. Two factors limit our sample. First, we only have daily inspection data for treatment districts, which include roughly 50 health inspectors. Of these inspectors, we drop fourteen who transferred into treatment districts taking over the phone of the previous inspector. Transfer records do not indicate the date of transfer, making it impossible to identify the period of smartphone data that correctly corresponds to these 14 inspectors.

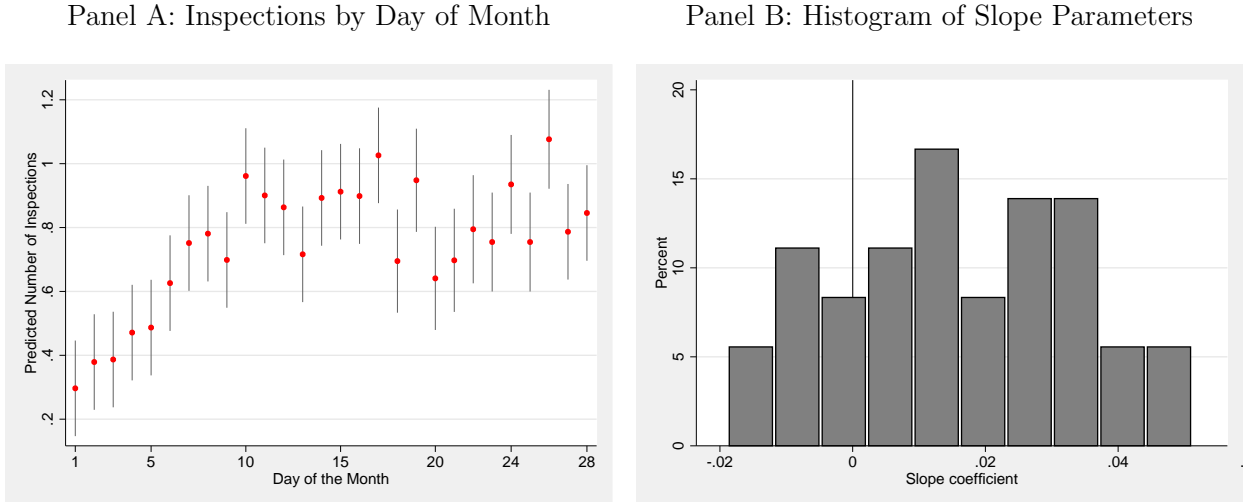


Figure 6: The Temporal Allocation of Inspections

*Notes:* Panel A plots the predicted number of inspections from a regression of inspections on dummies for each day of the month, and for each month, as well as a control for the number of facilities in the inspector’s jurisdiction. Panel B is a histogram of slope parameters obtained from estimating Equation (4) separately for each of the 36 vaccinators in our sample.

#### 4.1.6 Measuring Inspector Performance

We have two measures of job performance for health inspectors: (i) a dummy equal to one if the facility records an inspection in the two months prior to an unannounced visit; and (ii) the same proxy measure of collusion between doctors and health inspectors to falsify reports as described in Section 4.1.2. These measures were obtained during the same three independent and unannounced inspections of health clinics described in Section 4.1.2. Baseline performance measures for health inspectors are reported in Table A.1.

#### 4.1.7 Procrastination and Inspector Performance

As with our personality measures, we can correlate our proxy measure of the tendency to procrastinate with health inspector performance. In Table 1, we present results of a regression of health inspections on our estimated time slope coefficient. We see that health inspectors with larger time slope coefficients (reflecting a larger tendency to procrastinate) conduct fewer inspections, once you limit the sample to those inspectors with at least nine facilities in their jurisdiction (the 10th percentile in terms of health facilities per district

Table 1: Procrastination and Inspector Performance

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

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

across the sample). Specifically, we see that a one standard deviation increase in the procrastination measure is associated with a 6.7 percentage point decrease in the probability that an inspection was carried out in the last two months at a health clinic. This relationship may reflect a limitation on the number of inspections that can be carried out in a fixed period of time. Those who delay all of their inspections until the end of the month are not able to complete their monthly assignment.

#### 4.1.8 Personality and Inspector Performance

We examine how much the personalities of health inspectors predict their job performance in control districts (i.e., those under status quo incentives) in Figure 7. In Panel A, we consider the relation between personalities and whether an inspection was carried out in the last two months. In Panel B, we see that PSM traits are associated with less collusion, enough to

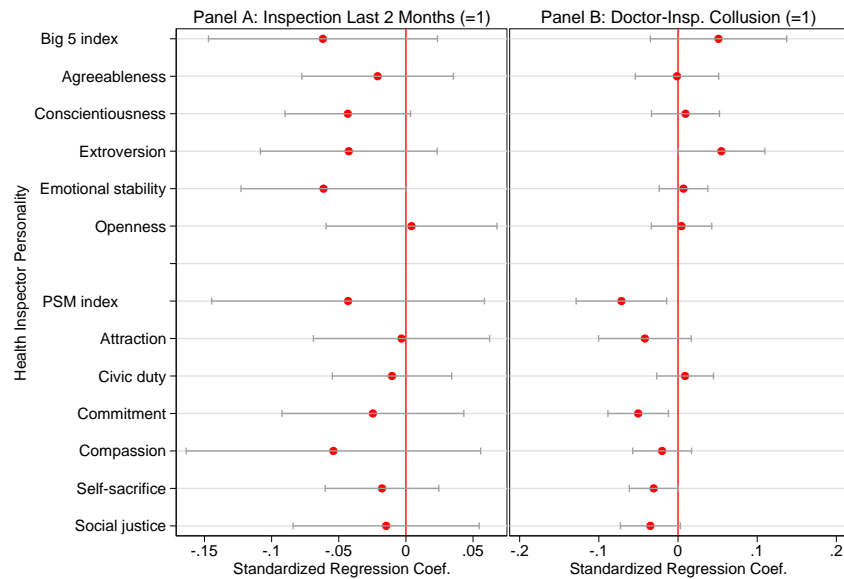


Figure 7: Personality and Performance: Health Inspectors

*Notes:* Each regression coefficient reported comes from a separate regression of the displayed performance measure on the indicated standardized health inspector personality measure. Error bars represent 95 percent confidence intervals. Standard errors are clustered at the clinic level. All regressions include Tehsil (county) and survey wave fixed effects. In all cases, personality measures are normalized to have mean zero and standard deviation of one in the sample, and thus the regression coefficients reported can be interpreted as the impact of a one standard deviation increase in a given personality trait or aggregate measure. The sample is restricted to control district clinics for which doctor personality data are available and a doctor is posted. Appendix Tables A.7 and A.8 provide corresponding regression tables.



distinguish the coefficient on the aggregate z-score from zero. In this case, health inspectors that score one standard deviation higher on aggregate PSM are about seven percentage points less likely to be identified as potentially colluding.<sup>43</sup>

## 4.2 Personalities and Treatment Response Heterogeneity

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

### 4.2.1 Evaluating the Smartphone Monitoring

Our experimental sample comprises all health facilities in the district of Punjab, which has a population of at least 85 million citizens. Tens of millions of public sector health users therefore were potentially affected by the program. As described above, we monitored a subsample of 850 clinics, drawn to be representative of facilities in the province, using independent and unannounced inspections.<sup>44</sup> 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 8 depicts control and treatment districts.<sup>45</sup>

---

<sup>43</sup>See Appendix Tables A.7 and A.8 for complete details on the results summarized in Figure 7. The estimates in Figure 7 indicate a negative relationship between both conscientiousness and emotional stability and the number of inspections. These coefficients both reflect  $p < 0.10$  and suggest that better traits are associated with worse performance. These coefficients are estimated only on the subsample of 298 clinics in control districts which have a doctor posted. In Appendix Tables A.9 and A.10, we find no evidence of a correlation on the full sample of 424 control facilities, indicating that inspectors with better traits are more likely to have inspected facilities *without* doctors posted. There is therefore some weak evidence that better inspectors substitute away from better facilities with a doctor posted toward more rural facilities without a doctor.

<sup>44</sup>These are the same clinics and inspections from the correlations presented in the previous section.

<sup>45</sup>Treatment is randomized at the district level. The intervention channels information about inspections to district health officials; a design randomizing treatment at an administrative unit beneath the district, say the tehsil, would very likely result in treatment affecting control units. The Department of Health also viewed sub-district randomization as not administratively feasible. Cluster randomization also allays some concerns about externalities generated by interactions between inspectors in the same district. All inspectors in a district are required to attend monthly meetings. While they typically have frequent interactions within districts, these relations are almost non-existent across districts.

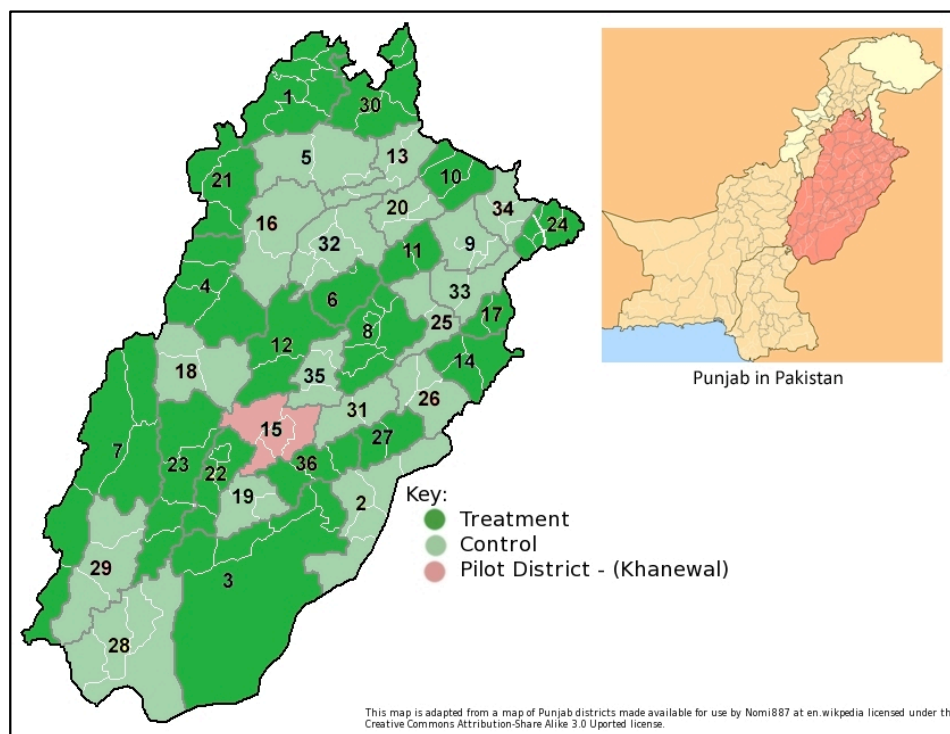


Figure 8: Treatment and Control Districts

#### 4.2.2 Personality and Treatment Response

We investigate whether impacts of the monitoring program are heterogeneous by the personality type of the inspector. Table 2 presents personality measures by treatment status for doctors and health inspectors. There is one significant difference in the balance table—treated health inspectors have slightly lower civic duty scores than those in control groups on average. This is plausibly due to sampling fluctuation as it is a fairly small difference and the only one among the 27 differences estimated.

We consider the effects of an increase in health inspector monitoring on their performance by inspector personality. Results are presented in Table 3.<sup>46</sup> We estimate regressions using the difference-in-difference specification

<sup>46</sup>Our other previous measure of performance, collusion between inspectors and doctors, cannot be studied in this context because the construction of collusion relies on data from our treatment districts' smartphone app. We have no information on health inspector-reported doctor attendance in the control districts of the Monitoring the Monitors experiment.

Table 2: Treatment Balance on Doctor and Health Inspector Personality

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

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

Table 3: Testing for Heterogeneous Impacts of Monitoring by Personality Type

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

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

$$Y_{dit} = \beta_0 + \beta_1 Trait_{di} + \beta_2 Treatment_{dit} + \beta_3 Treatment_{dit} \times Trait_i + \delta_t + \lambda_i + \varepsilon_{dit} \quad (5)$$

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

For health inspectors, there are heterogeneous effects of our experiment on the rate of health inspections. Health inspectors with a Big Five index one standard deviation above the mean, for example, exhibit a 35 percentage point higher treatment effect in terms of health inspections. With an unconditional mean inspection rate of 66 percent, inspectors with a z-score one standard deviation above the mean come very close to completing all of their inspections as a result of treatment. We decompose this effect in columns (5)-(9) and find that that it is being driven most strongly by emotional stability—the trait of being able to capably respond to new stressors and demands. Besides openness, all Big Five traits have positive and large coefficients. We also see some positive and similarly large effects of the PSM index, attraction, and compassion within the PSM traits, though only attraction is significant.<sup>47</sup>

Figure 9 presents nonparametric treatment effects of health inspector Big Five index across the distribution of inspectors according to the Big Five summary measure. We can see that the effect in Table 3 is primarily being driven by those health inspectors in the

---

<sup>47</sup>Note that to test for robustness in our effects to the small number of district clusters in our analysis, we have conducted Fisher exact tests for all results. In all cases, the estimated p-value is as at least as significant as from OLS. We have also separated the differential effects into our two post-treatment survey waves and find that the results sustain over time for as long as we were able to follow health clinics (roughly one year after treatment began). This is important because in Callen et al. (2013), we document that the overall treatment effects on health inspections do in fact fade by the second survey wave. Results available upon request.

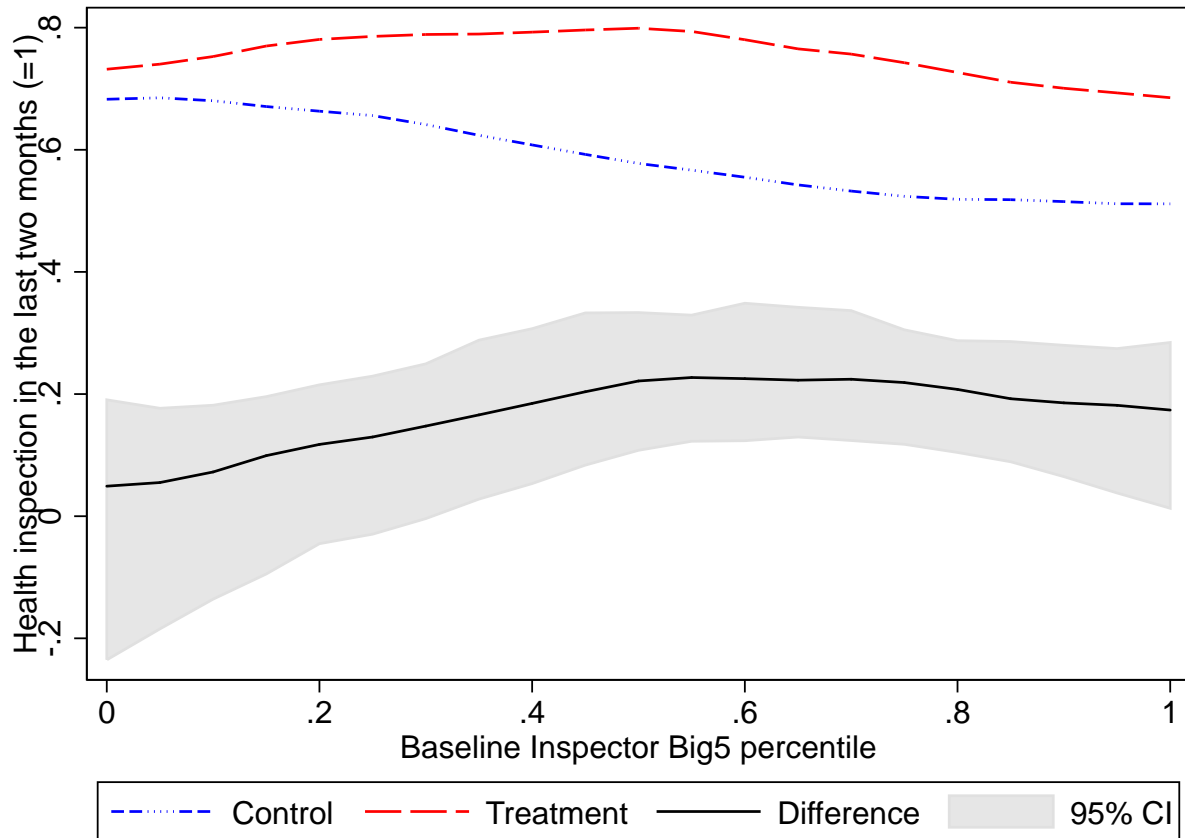


Figure 9: Nonparametric treatment effects

Figure 10: This figure plots a kernel-weighted local polynomial regression of whether a clinic had a health inspection in the last two months on every 5th percentile of baseline Big Five index separately for treatment and control districts, as well as the difference at each 5th percentile of baseline scores. The confidence intervals of the treatment effects are constructed by drawing 1,000 bootstrap samples of data that preserve the within-district correlation structure in the original data and plotting the 95 percent range for the treatment effect at each 5th percentile of baseline scores.

middle of the Big Five distribution. This fits the extended model presented in Section A.3 in which it is plausible that the effects of this intervention are localized to those inspectors in the middle of the distribution. See Appendix Figures A.1 and A.2 for nonparametric treatment effects trait-by-trait. While the location of the treatment effect peaks varies by trait, the overall shape is similar for specific traits.<sup>48</sup>

There are two more points to make about these experimental results. First, as you can see in Appendix Table A.12, personality does at least as much to predict the response to increased monitoring as all of the other covariates that we record for health inspectors. Completion of higher education is also a consistent predictor, but it predicts separately from personality. Second, these correlations are of a meaningful magnitude. Increased inspections may not lead to an overall increase in doctor attendance, but they generate information that is helpful in the case that a health inspector or more likely a senior health official *is* interested in enforcing attendance. We will see this directly in the next subsection.

### 4.3 Predicting Response to Information

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

#### 4.3.1 Information Experiment

The Monitoring the Monitors system aggregates data from health inspections and presents them to senior health officials in each district of Punjab on an online dashboard. This dashboard is only visible to these senior health officials as well as to the Secretary of Health

---

<sup>48</sup>Note that the point estimates in Figure 9 do not match those from Table 3. This is due to the fact that the regressions in the table include survey wave and clinic fixed effects.

Compliance Status Facility Status **Recent Visits** Indicators Time Trend Charts Photo Verification Map Change Password Logout

You are currently viewing **District Attock** (Please click to change view) [Print](#)

### Recent Facility Visits

■ Visits highlighted indicate significant staff absence.

BHU RHC THQ DHQ

Filter by Period   [Clear Filter](#)

Showing all entries

Displaying 1-30 of 734 result(s).

Go to page: < Previous **1** 2 3 4 5 6 7 8 9 10 Next >

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,	

Figure 11: Highlighting Underperforming Facilities to Test Mechanisms

for Punjab and the Director General of Health for Punjab. Figure 11 provides an example of a dashboard view visible to senior health officials.

To test whether senior health officials react to information about the absence of their subordinates, we directly manipulated the data on the dashboard to make certain underperforming facilities salient. This was achieved by highlighting in red, or “flagging” reports by inspectors that found three or more staff absent at a clinic.<sup>49</sup> This cutoff of three or more staff absences was set by our research team and was not communicated to any of the doctors, health inspectors, or senior health officials. We selected this cut-off based on the distribution of staff absence from baseline data. The peak of the distribution lies at two or three absent staff, suggesting that a cut-off at the center of this peak would yield the highest power to detect an effect of flagging in red.

Though the cutoff was purposefully arbitrary, our motivation for making absence data

<sup>49</sup>In Callen et al. (2013), we examine at length whether this manipulation affects subsequent doctor absence, finding consistent evidence that flagging facilities leads to decreased subsequent doctor absence.



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

### 4.3.2 Personality Predicts Response to Information

Appendix Table A.2 presents summary statistics for senior health officials in Punjab, which are similar in magnitude to summary statistics of both doctors and health inspectors. We examine whether manipulating attendance information affects subsequent doctor absence with the following specification

$$Absent\ Survey_{it} = \psi_0 + \psi_1 Trait_i + \psi_2 Flagged_{it-1} + \psi_3 Trait_i * Flagged_{it-1} + \delta_t + \eta_{it} \quad (6)$$

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

Facilities are flagged only if three or more staff members are absent. Consequently, if we restrict our sample to only facilities where, in the month prior to our unannounced visit, only two or three staff were absent, we can estimate the effect of flagging on a sample where the only difference might plausibly be whether the facility was flagged.<sup>50</sup>

Table 4 reports results from this test, limiting the sample to facilities with two or three staff absent during an inspection. Facilities flagged as underperforming to a senior official

---

<sup>50</sup>In Appendix Table A.13 we verify the drop in absence for people who score higher on the Big Five index is limited to right around the discontinuity, with a waning, though significant, effect in a slightly larger window.

Table 4: Tests of Heterogeneity in the Information Treatment by Senior Official Personality

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

*Notes:* This table tests for heterogeneity in the impact of providing information about underperforming clinics to senior officials by the personality types of the senior officials. Clinics were flagged as underperforming if three or more of the seven staff were absent in one or more health inspections of the clinic fifteen to forty-five days prior to an unannounced visit by our survey enumerators. All columns restrict the sample to those clinics where only two or three staff were absent (up to seven staff can be marked absent). In addition, the sample is limited to Monitoring the Monitor treatment districts due to the necessity of the web dashboard for flagging clinics. Column (1) reports un-interacted impacts of flagging. Columns (2) - (10) are further limited to clinics in districts for which senior health official personality data is available. The Big Five and PSM traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). All personality traits are then normalized across inspectors. The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively. Standard errors clustered at the clinic level reported in parentheses. All regressions include district and survey wave fixed effects and condition on a doctor being posted. *Levels of Significance:* \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

with a Big Five z-score one standard deviation above the mean subsequently experience an increase in doctor attendance that is 40 percentage points greater than a facility flagged to a senior official at the mean Big Five index.<sup>51</sup>

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

Appendix Table A.15 provides suggestive evidence that senior health officials with higher personality types stepped up the share of their time spent monitoring health facilities in response to dashboard flags. You can see senior health officials with a one standard deviation higher Big Five index increased the share of their time spent monitoring health facilities by 3.1 percentage points for each facility that was flagged in their district in the window prior to our collection of their time use information (wave three). The mean number of flags per district in this time-frame was 7.88, which translates to large increases in time spent monitoring by better personality types in response to flags. Although, this evidence is at best suggestive because it is based on seventeen observations.

The worry with the above results is that senior health officials might be substituting other work with increased monitoring of health facilities. The data suggest that senior health officials may have decreased their share of time spent on the lunch prayer break, on work related to monthly polio vaccination drives, and on ‘other work’ in response to flags.

---

<sup>51</sup>Note that in Table 4 we cannot reject the null hypothesis that the interaction term on the Big Five index is different than the uninteracted flagging effect. In Appendix Tables A.14, we show that when senior health officials’ are split into quartiles by Big Five index, we can significantly reject that those in the bottom and top quartile have the same flagging effect (with a substantial differential effect). We define the window during which a clinic can be flagged in red prior to one of our unannounced visits as 15 to 45 days before our visit. Senior health officials only looked at the web dashboard every week or two, so we would not expect an immediate response from flagging. However, if the window is made too long, virtually every facility will become flagged and we will lose variation. The p-values of the significance of the coefficient on the Big Five index and PSM index for a wide range of windows are reported in Appendix Figures A.3 and A.4. These figures also indicate that we have not selected the window most favorable for our result.

Unfortunately, these effects are not significant individually.<sup>52</sup>

As with the correlational and experimental results above, we show that personality is a better predictor of the response to information than other important covariates for senior health officials. See Appendix Table A.16 for these results.

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

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

## 4.4 Summary of Results

This paper performs several tests of associations between personalities and objective performance measures for public health workers at three different levels of the bureaucracy in Punjab, Pakistan. These results are summarized in Table 5.<sup>53</sup>

We run regressions reflecting seven separate tests for eleven different traits, for a total

---

<sup>52</sup>Category-by-category time use tables available by request.

<sup>53</sup>Appendix Table A.17 presents the same table but with standard errors adjusted using the False Discovery Rate procedure in Benjamini and Hochberg (1995).

Table 5: Results Summary

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

*Notes:* This table provides a summary of the results available in Figures 4 and 7 and Tables 3 and 4. + (++) indicates a positive relationship significant at the 10 percent (5 percent) level and - (--) indicates a negative relationship significant at the 10 percent (5 percent) level.

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

## 5 Conclusion

Governments, like any organization, are made of people with potentially stark interpersonal differences. We find that these differences are useful in predicting performance. In addition, our data also allow us to examine the pattern of relationships between specific traits and performance.

Five of the seven tests linking conscientiousness—the personality trait of being responsible, hardworking, and desiring to do tasks well—to performance are statistically significant with at least 90 percent confidence and always in the predicted direction. This trait appears to be important for doctors, inspectors, and senior officials alike. Similarly, four of

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

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

## References

- Almlund, Mathilde, Angela Lee Duckworth, James J Heckman, and Tim D Kautz**, “Personality psychology and economics,” Technical Report, National Bureau of Economic Research 2011.
- Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack**, “No margin, no mission? A Field Experiment on Incentives for Public Service Delivery,” *Journal of Public Economics*, Forthcoming, 8834.
- , – , and **Scott S Lee**, “Do-gooders and go-getters: career incentives, selection, and performance in public service delivery,” 2014.
- Augenblick, Ned, Muriel Niederle, and Charles Sprenger**, “Working Over Time: Dynamic Inconsistency in Real Effort Tasks,” *The Quarterly Journal of Economics*, Forthcoming.
- Bandiera, Oriana, Andrea Prat, and Tommaso Valletti**, “Active and Passive Waste in Government Spending: Evidence from a Policy Experiment,” *American Economic Review*, 2009, 99 (4), 1278–1308.
- Barrick, Murray R. and Michael K. Mount**, “The Big Five Personality Dimensions and Job Performance: A Meta-Analysis,” *Personnel Psychology*, 1991, 44 (1), 1–26.
- Bazerman, Max and Don A Moore**, *Judgment in Managerial Decision Making, 8th Edition*, Wiley & Sons, 2012.
- Benabou, Roland and Jean Tirole**, “Intrinsic and extrinsic motivation,” *The Review of Economic Studies*, 2003, 70 (3), 489–520.
- Benjamini, Yoav and Yosef Hochberg**, “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, pp. 289–300.
- Bertrand, Marianne and Antoinette Schoar**, “Managing with Style: The Effect of Managers on Firm Policies,” *Quarterly Journal of Economics*, 2003, CXVIII, 1169–1208.
- Besley, Timothy and Maitreesh Ghatak**, “Competition and incentives with motivated agents,” *American Economic Review*, 2005, 95 (3), 616–636.
- Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan**, “Reducing crime and violence: Experimental evidence on adult noncognitive investments in Liberia,” 2015. Working paper.
- Borghans, Lex, Angela Lee Duckworth, James J. Heckman, and Bas ter Weel**, “The Economics and Psychology of Personality Traits,” *The Journal of Human Resources*, 2008, XLIII (4), 973–1059.
- Bowles, Samuel, Herbert Gintis, and Melissa Osborne**, “The determinants of earnings: A behavioral approach,” *Journal of Economic Literature*, 2001, pp. 1137–1176.

- Callen, Michael, Saad Gulzar, Ali Hasanain, and Yasir Khan**, “The Political Economy of Public Employee Absence: Experimental Evidence from Pakistan,” 2013.
- Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F Halsey Rogers**, “Missing in action: teacher and health worker absence in developing countries,” *The Journal of Economic Perspectives*, 2006, *20* (1), 91–116.
- Chetty, Raj, John N Friedman, and Jonah E Rockoff**, “Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates,” *American Economic Review*, 2014.
- , – , and – , “Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood,” *American Economic Review*, 2014.
- Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi**, “Strengthening State Capabilities: The Role of Financial Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service,” *Quarterly Journal of Economics*, 2013, *forthcoming*.
- Deaton, Angus**, “Instruments, Randomization, and Learning about Development,” *Journal of Economic Literature*, 2010, *48* (2), 424–55.
- Gatewood, Robert, Hubert Feild, and Murray Barrick**, *Human resource selection*, Cengage Learning, 2010.
- Groth-Marnat, Gary**, *Handbook of psychological assessment*, John Wiley & Sons, 2009.
- Hanushek, Eric A**, “The economic value of higher teacher quality,” *Economics of Education Review*, 2011, *30* (3), 466–479.
- and **Steven G Rivkin**, “The distribution of teacher quality and implications for policy,” *Annu. Rev. Econ.*, 2012, *4* (1), 131–157.
- Heckman, James J.**, “Integrating Personality Psychology into Economics,” 2011, (NBER WP #17378).
- , **Jora Stixrud, and Sergio Urzua**, “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior,” *Journal of Labor Economics*, 2006, *24* (3), 411–482.
- 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**, “Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues,” in “Handbook of personality: Theory and research,” The Guilford Press, 2008, chapter 4.



- Johnson, W. Bruce, Robert Magee, Nandu Nagarajan, and Harry Newman**, “An Analysis of the Stock Price Reaction to Sudden Executive Deaths,” *Journal of Accounting and Economics*, 1985, 7, 151–174.
- Jones, Benjamin F. and Benjamin A. Olken**, “Do Leaders Matter? National Leadership and Growth Since World War II,” *Quarterly Journal of Economics*, 2005, 120 (3), 835–864.
- Kaplan, Robert M. and Dennis P. Saccuzzo**, *Psychological Testing: Principles, Applications, and Issues*, Pacific Grove, Calif.: Brooks/Cole Pub. Co., 1997.
- Kautz, Tim D, James J. Heckman, Ron Diris, Bas ter Weel, and Lex Borghans**, “Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success,” 2014, (NBER WP #20749).
- Klinger, Bailey, Asim Ijaz Khwaja, and Carlos del Carpio**, *Enterprising Psychometrics and Poverty Reduction*, Springer, 2013.
- Kuhn, Michael A**, “Curing the Calorie Crunch: The Effect of EBT on Household Present-Bias,” 2013.
- 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.
- Miguel, Edward, C Camerer, K Casey, J Cohen, KM Esterling, A Gerber, R Glennerster, DP Green, M Humphreys, G Imbens et al.**, “Promoting transparency in social science research,” *Science*, 2014, 343 (6166), 30–31.
- National Institute of Population Studies**, *Pakistan Demographic and Health Survey 2012-13*, National Institute of Population Studies, 2013.
- Nyhus, Ellen K. and Empar Pons**, “The Effects of Personality on Earnings,” *Journal of Economic Psychology*, 2005, 26 (3), 363–384.
- Olken, Benjamin A. and Rohini Pande**, “Corruption in Developing Countries,” *Annual Review of Economics*, 2012, 4, 479–509.
- Perry, James L.**, “Measuring Public Service Motivation: An Assessment of Construct Reliability and Validity,” *Journal of Public Administration Research and Theory*, 1996, 6 (1), 5–22.
- **and Lois Recascino Wise**, “The Motivational Bases of Public Service,” *Public Administration Review*, 1990, 50, 367–73.
- Petrovsky, Nicolai**, “Does Public Service Motivation Predict Higher Public Service Performance? A Research Synthesis,” 2009.
- Rasul, Imran and Daniel Rogger**, “Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service,” 2014.

- Reinikka, Ritva and Jakob Svensson**, “Local Capture: Evidence from a Central Government Transfer Program in Uganda,” *The Quarterly Journal of Economics*, 2004, 119 (2), 679–705.
- Roberts, Brent W.**, “Back to the Future: Personality and Assessment and Personality Development,” *Journal of Research in Personality*, 2009, 43 (2), 137–145.
- , **Kate E. Walton, and Wolfgang Viechtbauer**, “Patterns of Mean-Level Change in Personality Traits across the Life Course: A Meta-Analysis of Longitudinal Studies,” *Psychological Bulletin*, 2006, 132 (1), 1–25.
- Salgado, Jesus F.**, “The Five Factor Model of Personality and Job Performance in the The Five Factor Model of Personality and Job Performance in the European Community,” *Journal of Applied Psychology*, 1997, 82 (1), 30–43.
- Schmidt, Frank L and John E Hunter**, “The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings.,” *Psychological bulletin*, 1998, 124 (2), 262.
- Shapiro, Jesse M.**, “Is there a daily discount rate? Evidence from the food stamp nutrition cycle,” *Journal of Public Economics*, 2005, 89 (2), 303–325.
- Staiger, Douglas O and Jonah E Rockoff**, “Searching for effective teachers with imperfect information,” *The Journal of Economic Perspectives*, 2010, 24 (3), 97–117.
- Wild, Lena, Vikki Chambers, Maia King, and Daniel Harris**, “Common Constraints and Incentive Problems in Service Delivery,” Technical Report, Overseas Development Institute 2012.
- World Bank**, *World Development Report 2004: Making Services Work for the Poor*, World Bank, 2004.

## A Appendix - For Online Publication Only:

### A.1 Appendix Tables

Table A.1: Doctor and Health Inspector Personality Summary Statistics (Control Districts)

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

*Notes:* Sample for Panel A: doctors in control districts that completed the personalities survey module, given in waves 2 and 3 and during a special follow-up round. Sample for Panel B: health inspectors in control districts that completed the personalities survey module. Doctors and inspectors were only asked to complete the module once. The Big Five and PSM traits are each mean responses to statements that represent the trait on a five point likert scale, in which 1 corresponds to disagree strongly, 2 to disagree a little, 3 to neutral, 4 to agree a little, and 5 to agree strongly. Likert responses are given the same direction (5 always being more agreeable, for example, never less). The Big Five and PSM Indices are z-score averages of the five and six traits within the Big Five and PSM respectively. Performance and collusion samples are clinic-wave observations in control districts across waves 1 through 3, where doctors are posted. Collusion is a dummy variable coded as 1 when a doctor is reported absent in both survey waves 2 and 3 but is reported as present by health inspectors during every visit between the launch of the program and present (up to 50 visits).

Table A.2: Senior Health Official Personality Summary Statistics (Control Districts)

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

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

Table A.3: Doctor Personality and Doctor Attendance

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

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

Table A.4: Doctor Personality and Estimated Doctor-inspector Collusion

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

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

Table A.5: Doctor Personality Measure Predictions Compared to Other Covariates

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

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



Table A.6: Doctor Attendance and Health Service Provision (Control Districts)

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

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

Table A.7: Health Inspector Personality and Inspections

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

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

Table A.8: Health Inspector Personality and Estimated Doctor-inspector Collusion

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

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

Table A.9: Health Inspector Personality and Inspections—Full Sample

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

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

Table A.10: Health Inspector Personality and Estimated Doctor-inspector Collusion—Full Sample

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

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

Table A.11: Personalities and Health Inspections—Experimental Evidence, Unconditional on Doctor Being Posted

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

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

Table A.12: Inspector Personality Measure Experimental Results Compared to Other Covariates

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

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

Table A.13: Differential Clinic Flagging Effects by Senior Health Official Personality, Robustness to Cutoff

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

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



Table A.14: Differential Clinic Flagging Effects by Senior Health Official Personality, Semi-parametric

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

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

Table A.15: Differential Senior Health Official Time Use by Personality

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

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

Table A.16: Differential Clinic Flagging Effects by Senior Health Official Personality Compared to Other Covariates

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

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

Table A.17: Results Summary

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

*Notes:* +/- indicates sign of coefficients from regression reported in Figures 4 and 7 and Tables 3 and 4. For information on each specific hypothesis test, see these. The number of + or -'s represents the p-value from the reported hypothesis test (one corresponds to a p-value < 0.1, two to a p-value < 0.05). All p-values are corrected for multiple hypothesis testing using the False Discovery Rate procedure outlined in Benjamini and Hochberg (1995). P-values for the 14 index hypotheses are corrected separately than the p-values for the 55 trait hypotheses.

## A.2 Appendix Figures

Figure A.1: Health Inspector Non-parametric Heterogeneous Effects, Trait-by-Trait, Big Five

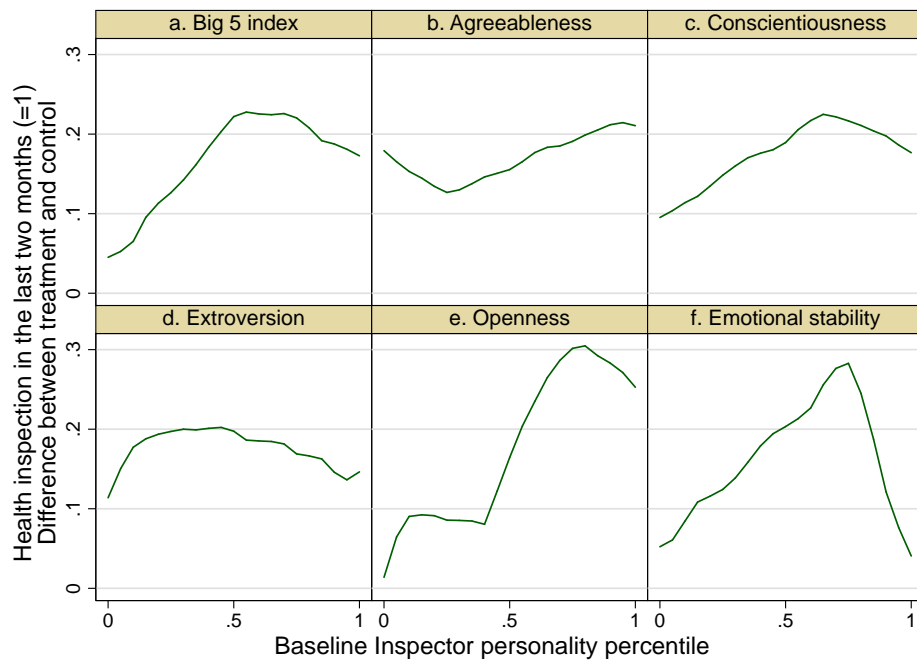


Figure A.2: Health Inspector Non-parametric Heterogeneous Effects, Trait-by-Trait, PSM

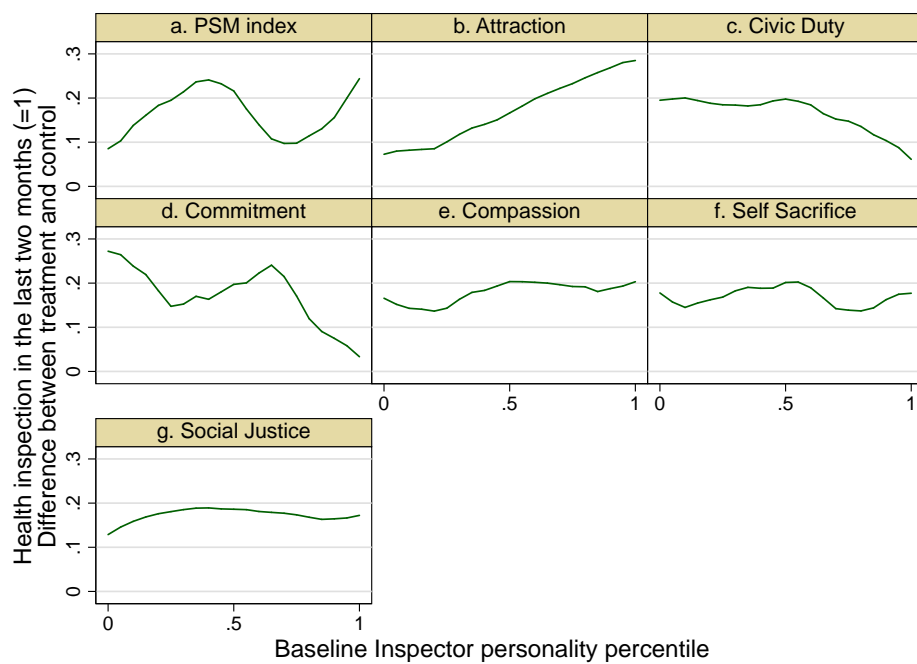


Figure A.3: Robustness to Different Windows for Flagging- Big Five Index

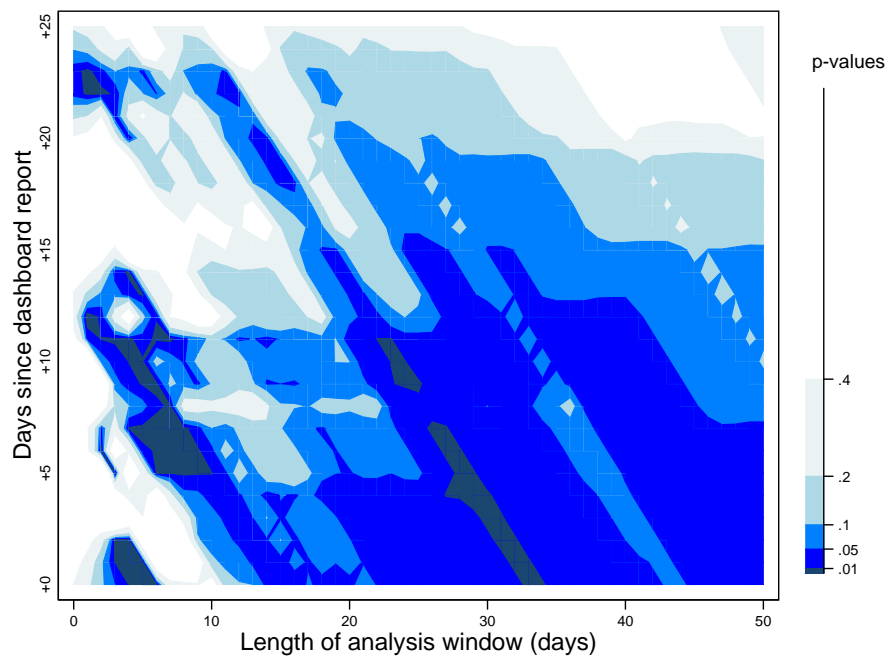
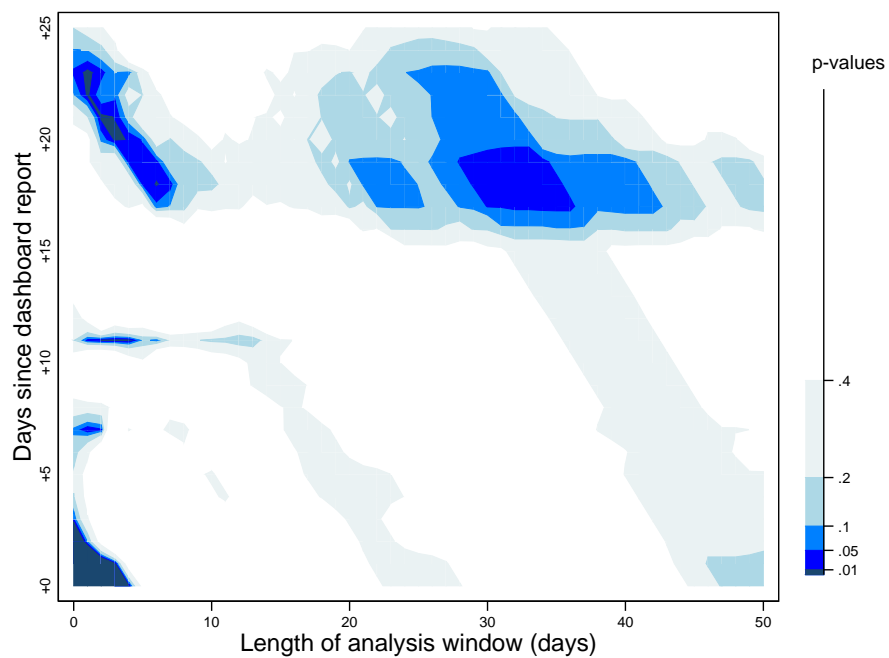


Figure A.4: Robustness to Different Windows for Flagging- PSM Index



### A.3 Model extension

As with the model in Section 2, let our personality measures represent a worker's type,  $\theta$ , with cumulative distribution  $F(\theta)$ . Let performance be the binary decision that a doctor or health inspector makes of whether to show up to work or to shirk. If a worker chooses to work, he receives a fixed salary of  $W$  and incurs a cost of effort of  $\lambda(\theta)$ . If a worker chooses to shirk, he exerts no effort and receives the fixed salary with probability  $1-p$  and an arbitrarily small punishment  $c$  with probability  $p$  as well as an outside option of  $Q$ . However, let us now assume that the outside option is a function of  $\theta$ . Thus we have the following updated indifference condition:

$$W - \lambda(\theta) = (1 - p)W - pc + Q(\theta) \quad (7)$$

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

Given these assumptions, we can plot the net payoff to working versus the net payoff to shirking before and after an increase in  $p$  under various scenarios. Both Figure A.5 and A.6 show a case when the  $\lambda(\theta)$  function is linear and the  $Q(\theta)$  function is convex in  $\theta$ .<sup>54</sup>

These figures allow us to make several important points. First, we can see that in both figures an increase in incentives to work induces a range of workers in the middle of the personality type distribution to work. Second, we can see that in the second figure, before an increase in  $p$  no one chooses to work. This highlights that the existence of a relationship between performance and personality type is subject to the outside option for some personality types being sufficiently low. More generally, the difference between the

---

<sup>54</sup>Note that the case when both functions are linear is very unlikely to be accurate, while the case when both functions are strictly convex, while likely more accurate, does not lead to any additional intuition (both presented cases would hold so long as the  $\lambda(\theta)$  function has less curvature than the  $Q(\theta)$  function over the relevant range).



two figures highlights the ambiguity in correlation between performance and personality type under a fixed  $p$ . In the first figure, all workers above a certain marginal worker will choose to work, with the marginal worker shifting to the left after  $p$  is increased. This would create a positive correlation between personality type and working under the status quo and a positive correlation between personality type and responding to an increase in  $p$  by switching from shirking to working. Where as in the second figure, the gains to the outside option for the highest personality types overcome the gains to those types for working even after  $p$  is increased sufficient to induce some personality types to work, causing the best personality types to join the worst personality types in shirking. This would lead to an ambiguous correlation between personality type and working.

Figure A.5: Effect of an Increase in Detection Probability on the Decision to Work or Shirk

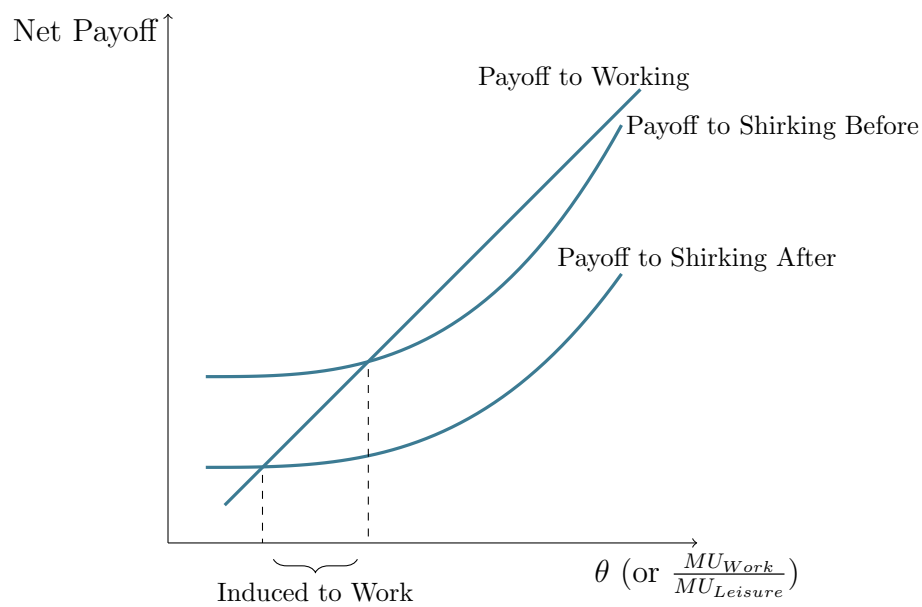
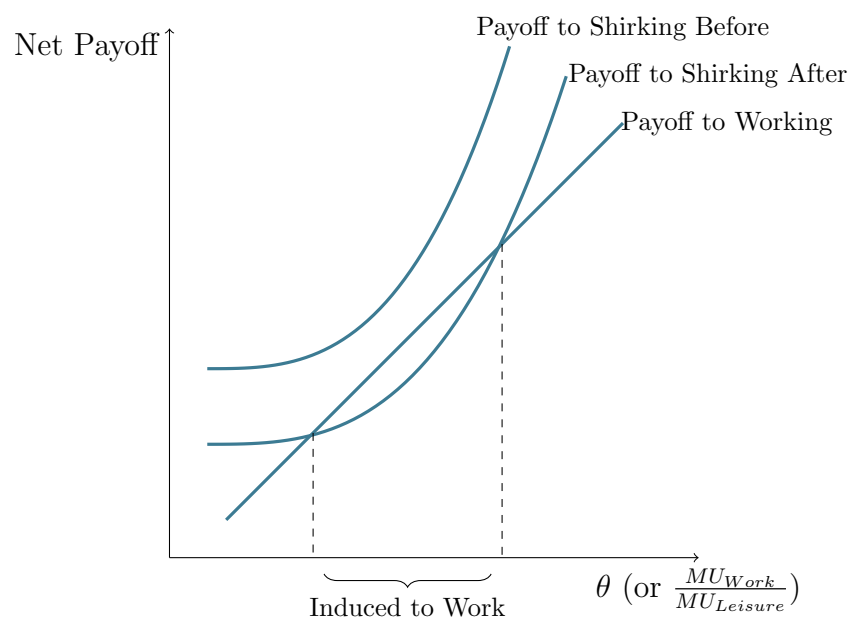


Figure A.6: Effect of an Increase in Detection Probability on the Decision to Work or Shirk



## A.4 The Effects of Improved Doctor Personality on Health Service Provision

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

On average, the bottom 25% of doctors in our sample in control groups have a normalized Big Five index of -0.95 standard deviations. The mean doctor in the control groups has a normalized Big Five index of 0.05 standard deviations. Thus replacing the average bottom quartile doctor with an average doctor would raise doctor Big Five index by 1 standard deviation. This is then associated with 3.7 percentage point increase in doctor attendance respectively (combining the average Big Five index increase with the coefficient from Appendix Table A.3). Now combining with our our results from Appendix Table A.6, a 3.7 percentage point increase in doctor attendance is associated with an average increase of 7.45 outpatients seen per month. There are approximately 2496 total clinics in Punjab. Thus if we were to replace the bottom 25% of doctors in our sample with average doctors in terms of Big Five index, we would expect to increase total outpatients seen per month in Punjab by 4646.46 visits per month. This is a very meaningful number. And this is only half of what we would expect if we did the same exercise in terms of conscientiousness, the personality trait with the strongest correlational result for doctors. In summary:

Table A.18: Doctor Attendance and Health Service Provision Appendix Exercise

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

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

## A.5 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 by 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"