Jobs for Sale: Corruption and Misallocation in Hiring

Jeffrey Weaver*

July 6, 2020

Abstract

Corruption in government hiring is common in developing countries and thought to worsen service delivery outcomes. I collect original data on hiring in a health bureaucracy and find hires pay bribes averaging 17 months of salary. Despite this, the quality of corrupt hires is high and compares favorably even to merit-based hires. However, there is meaningful heterogeneity, with corruption causing poor quality hires in a quarter of cases. Exploiting variation across jobs, I show why corruption can lead to relatively efficient outcomes and how the extent of misallocation from corruption is explained by a simple set of economic parameters. JEL codes: O01, D73, M05

*I thank Chris Udry, Mark Rosenzweig, Nick Ryan, Ebonya Washington, Vittorio Bassi, Alex Cohen, Saby Das, Lauren Falcao Bergquist, Sharat Ganapati, Robert Gonzalez, Dean Karlan, Dan Keniston, Matt Lowe, Karthik Muralidharan, Paul Niehaus, Sam Norris, Rossa O’Keefe-O’Donovan, Paulina Oliva, Ben Ollken, Matt Pecenco, Nancy Qian, Santiago Saavedra, Gabriella Santangelo, Meredith Startz, Sandip Sukhtankar, Edoardo Teso, Jaya Wen, and seminar audiences at UC-Berkeley, Berkeley-Haas, Center for Global Development, EconCon, EMCON, NBER Summer Institute (Development & Personnel), NEUDC, Ohio State, UC-San Diego, USC, and Yale for helpful comments, as well as Scott Lee for sharing materials. All fieldwork and analysis was conducted while the author was supported by a NSF Graduate Research Fellowship. Funding for data collection was provided by a Yale Economic Growth Center/Sylff Foundation Research Grant. Yale IRB: 1408014406. All views expressed are those of the author and do not necessarily reflect the opinions of any of the funding organizations.

Email: jbweaver@usc.edu
I Introduction

The recruitment of public sector workers is an important determinant of state capacity and service delivery outcomes (Dal Bó et al., 2013; Finan et al., 2017; Best et al., 2019). Previous studies have theorized that corruption in the allocation of government jobs, in which hiring decisions are made on the basis of bribes or connections, will have negative consequences for service delivery and may be one of the root causes of bureaucratic inefficiency (e.g. Wade 1982; Shleifer and Vishny 1993; Muralidharan 2015; Sukhtankar and Vaishnav 2015). Such corruption is thought to be widespread in the developing world: for example, Kristiansen and Ramli (2006) interviewed 60 Indonesian civil servants and found that all had paid a bribe to be hired, while former Russian Prime Minister Dmitry Medvedev has publicly acknowledged that most government jobs in Russia can be purchased (NewsRu, 2008). However, the difficulty of collecting data on corruption in hiring has resulted in little empirical evidence on its effects.

Theoretical models provide ambiguous predictions on the impact of corruption. Some papers have argued that corruption may be economically efficient in allocating scarce goods to individuals who value them most highly (Leff, 1964; Beck and Maher, 1986; Lien, 1986; Shimer, 1999). On the other hand, corruption may allocate valuable resources such as jobs to the corrupt or wealthy rather than the socially efficient recipients (Krueger, 1974; Esteban and Ray, 2006). This theoretical ambiguity generates important empirical questions: can corruption lead to efficient outcomes, and if so, what factors determine whether corruption leads to efficient or inefficient outcomes?

In this paper, I collect original data from a corrupt government hiring process for jobs as supervisors of community health workers (CHWs). There are over five million CHWs globally, and their work has been shown to be an important determinant of public health outcomes (Deserranno, 2019; Ashraf et al., 2020). The study area is in a rural portion of a large developing country and is served by over a thousand CHWs.\(^1\) These CHWs had worked in their positions for an average of eight years prior to this hiring, but the supervisor position did not previously exist. For the hiring process, current CHWs were grouped into geographically-based clusters of 15 to 25 CHWs, and one supervisor was hired for each cluster. Only the CHWs within a given cluster were eligible to apply for that cluster’s supervisor position, a feature which will prove important for the empirical strategy. Hiring processes like this one, with a small number of desirable positions and many similarly credentialed applicants, are common in modern developing country bureaucracies.

The first part of the paper shows that bribes were the main determinant of hiring decisions. I collect and cross-validate data on bribe payments from key informants involved in the hiring process, including hires and unsuccessful applicants. All hires made bribe payments, and these payments averaged 17 months of salary as a supervisor.\(^2\) Although around half of unsuccessful

\(^1\) Due to human subjects concerns about respondents and data collectors being identified and retaliated against, the identity of the country was removed from the paper. Exposing corruption can be costly, with many cases of anti-corruption activists losing their jobs or being murdered. This precaution is also taken by Cole and Tran (2011), which examine a firm’s records of bribe payments, and De Janvry et al. (2015) due to a non-disclosure agreement.

\(^2\) Being hired as a supervisor resulted in a roughly 40% increase in pay, making it a reasonable financial investment. Journalists have documented similarly high prices for government jobs in other locations, such as Afghani border
applicants report making a bribe offer to the hiring committee, only those hired actually made payments. Taking the characteristics and bribe offers of both hires and unsuccessful applicants, I use the hiring agent’s revealed preference among the applicants for each job to characterize how they made hiring decisions. Bribe size is the primary determinant of who is hired, while political connections and the education of applicants play a minor role. The data are inconsistent with other models of corrupt hiring processes, such as meritocratic hiring with ex post extortion for bribes.

The second section of the paper studies the extent to which hiring decisions deviated from the first-best case. Since only current CHWs were eligible to apply and I collect information on all CHWs, I can compare hires to the universe of individuals who were eligible or applied for these jobs. Hires appear to be high quality, as they are better than non-hires across most possible quality measures, including cognitive ability and past performance as a CHW. However, the policy-relevant counterfactual is how hires compare to those who would have been hired under merit-based systems.

For this comparison, I proceed in two steps: first determining who would have been hired under counterfactual systems and then comparing their quality to that of actual hires. I focus on counterfactual hires under three merit-based systems: 1) a knowledge-based test; 2) a test of problem solving ability; and 3) the rules that were supposed to be used for hiring decisions in this context. Identifying these hires requires identifying which of the potential applicants would have applied and been hired under each system, as application decisions may differ across systems (e.g. some individuals may not have applied under the corrupt system due to a distaste for bribery, but would apply under a merit-based system). I use a simple entry model to estimate application decisions and validate the model with tests based on actual application decisions.

I then compare actual and counterfactual hires using a set of characteristics related to supervisor performance. Taking a 20 month panel data set on delivery of health services by CHWs, I estimate which supervisor characteristics are related to improvements in service delivery by the CHWs they supervise and construct an index of these characteristics, which I label “supervisor performance index” (SPI). Since these characteristics are observed for all CHWs, I can compare the actual and counterfactual hires on the index. The quality of actual hires compares favorably to that of counterfactual hires. The value of SPI among actual hires is on average at 90% of the “first-best” case, while hires under the health knowledge test are at only 83%. Those who would have been hired under the intended hiring rules were weakly worse than actual hires (88% of first-best case), while hires under the test of problem solving were slightly better than the actual hires (93.5% of first best case). Translating these differences into service delivery outcomes and monetary terms using the value of a statistical life, the social welfare implications of differences in supervisor quality is large and far exceeds the value of bribes. However, these average effects mask considerable heterogeneity: for example, 26% of actual hires were in the bottom half of the SPI distribution of officials - $15,000 (Walsh, 2014); Chinese train attendants - $14,897 (Moore, 2013); Indian police recruits - $4,000 to $6,000 (Jauregui, 2007); and Romanian doctors - $20,000 to $40,000 (Stancu, 2014).

The “first-best” is if the highest SPI candidate were hired. To calculate the percent of the first-best case, I take the value of SPI for each hire, subtract the lowest SPI of a candidate eligible to apply for the position, and divide by the difference between maximum and minimum SPI of potential candidates for the position. This ranges between 0 and 1 for the lowest and highest SPI candidate respectively being selected.
applicants for their position. This generates two puzzles: first, why is the average quality of hires in a corrupt system relatively high, and second, what causes variation in the degree of misallocation from corruption?

The final section of the paper addresses these questions by demonstrating a mechanism that determines when corruption leads to efficient outcomes. Public sector jobs are one of a large class of goods that bureaucrats may allocate corruptly. Some theoretical work has argued that such allocations will be efficient (Leff, 1964; Beck and Maher, 1986; Lien, 1986), while others have argued that corrupt allocations will be inefficient (Shleifer and Vishny, 1993; Esteban and Ray, 2006). An alternative possibility is that the consequences of corruption are not uniformly bad or good, but varies depending on features of the institutional context (Banerjee et al., 2013). I test a model in which the efficiency consequences of corruption depend on two economic parameters: in the case of a corrupt allocation of jobs, these are the correlations between how well candidates would perform in the job and their wealth and valuation of the job.

This context offers an ideal setting to empirically test this model. Each supervisor position had a fixed and externally determined pool of eligible candidates: only current CHWs could apply, and they were restricted to applying to the one position that their work area falls under. This generates variation across positions in the correlations of quality with wealth and valuation in their candidate pool. Using a “leave-out” version of SPI to measure quality, the correlation of quality and wealth among the pools of eligible candidates varies substantially: it is positive and greater than 0.2 for 64% of positions and negative for 26% of positions. Leveraging variation in the composition of candidate pools that is plausibly exogenous to pre-existing service delivery outcomes, I test whether outcomes are relatively efficient (i.e. higher quality hires, better service delivery outcomes) when the correlation of wealth and quality is positive, and less efficient when the correlation is negative.

Consistent with the hypothesis, I find hires are better quality and service delivery outcomes improve by more after the hiring of supervisors in clusters where the correlation of wealth and quality is more positive.4 This explains both puzzles from the second section of the paper: outcomes were on average close to first-best due to the strong positive correlations between SPI, valuation and wealth across all CHWs in the region. However, there is heterogeneity because for those jobs in which these correlations were negative, allocations are relatively inefficient.

Government allocations of scarce resources such as jobs are characterized by multiple market failures. Bureaucrats typically lack incentives to make socially optimal allocations or design a screening mechanism that can do so; they may also lack information on how to optimally screen. Given these frictions, this paper shows that bribery can be a second-best screening tool, but that the efficiency consequences of allocation by bribes will depend on a simple set of economic parameters. These findings can also be extended to other contexts in which governments allocate scarce goods and for considering where anti-corruption efforts may have the largest impact.

---

4The key identifying assumption is that this correlation impacts service delivery through the hiring of supervisors. This is supported by numerous tests: the correlation of wealth and quality is not related to cluster characteristics (e.g. average wealth or quality of CHWs), or pre-existing levels or trends of service delivery outcomes, i.e. it only becomes relevant for service delivery when supervisors are hired.
This paper contributes to a number of distinct literatures. First, this paper is related to work on recruitment and hiring of public sector workers (e.g. Dal Bó et al. 2013; Hanna and Wang 2017; Finan et al. 2017). Along with the contemporary work of Colonnelli et al. (2019), Ornaghi (2019) and Xu (2018), which all study patronage-based rather than bribery-based hiring, this is one of the first papers to directly measure the consequences of non-meritocratic hiring.\(^5\) In contrast to those papers, other work on selection based on connections (e.g. Fisman et al., 2018), and most theoretical discussions of corruption-based hiring (Shleifer and Vishny, 1993; Muralidharan, 2015; Sukhtankar and Vaishnav, 2015), I find that non-meritocratic hiring need not distort the efficient allocation of jobs.

Second, the paper speaks to broader questions of how corruption affects allocative efficiency. The concept of efficient corruption is well known theoretically (Leff, 1964; Beck and Maher, 1986; Lui, 1985), but there is minimal empirical evidence in practice.\(^6\) The vast majority of the empirical literature on how corruption affects economic outcomes has found negative consequences from embezzlement (e.g. Olken 2007; Ferraz et al. 2012; Muralidharan et al. 2017b), extortion (e.g. Sequeira and Djankov, 2014), exchange of favors (e.g. Bertrand et al. 2007; Duflo et al. 2013), and corruption in allocation of in-kind transfers (Niehaus et al., 2013). This paper departs from existing work in demonstrating the heterogeneous effects of corruption and documenting the economic parameters that underlie this heterogeneity. This framework can be applied to many other types of public goods and is relevant in considering where cracking down on corruption can have the highest returns.

Third, this paper contributes to the literature documenting how corrupt markets work (Olken and Barron, 2009; Burgess et al., 2012). Corruption takes many forms with different underlying economic structures, and understanding the nature of the market is relevant in the design of anti-corruption policy. Existing documentation of corrupt markets comes primarily from settings such as the exchange of money for favors (e.g. Bertrand et al. 2007; Duflo et al. 2013), kick-backs in bidding for contracts (Cole and Tran, 2011), and embezzlement (e.g. Olken 2007; Niehaus and Sukhtankar 2013). In the case of markets for jobs, nearly all empirical papers have focused solely on the role of patronage (Iyer and Mani, 2011; Fafchamps and Labonne, 2017; Colonnelli et al., 2019) aside from the more qualitative accounts of Wade (1982) and Wade (1985). This paper uses detailed data to understand the role of bribery in non-meritocratic hiring, including how bribery and patronage concerns interact. I find that patronage is a secondary consideration to bribes in this context, highlighting the importance of studying factors in addition to patronage in understanding developing country bureaucracies.

The remainder of the paper is organized as follows. Section 2 details the context, data collection,

\(^5\)There is a separate literature on the historical sale of offices in medieval Europe (e.g. Allen, 1998), colonial governorships (Guardado, 2018), and tax farming (e.g. Johnson and Koyama, 2014; White, 2004). These differ from the context of this paper since they are about the sale of offices by the state. In the contemporary context, sale by unauthorized actors is much more common, with potentially more negative consequences.

\(^6\)I know of only two papers finding evidence of efficient corruption, both due to different underlying mechanisms than this paper. Dreher and Gassebner (2013) find evidence of “greasing the wheels” (Lui, 1985), while in Sukhtankar (2015), there is a reallocation of licenses to efficient firms after an initially corrupt allocation.
and market for jobs. Section 3 compares the actual, corrupt hires to those who would have been hired under merit-based processes. Section 4 discusses the heterogeneous effects of corruption and demonstrates the mechanism underlying this heterogeneity. Section 5 concludes.

II Background and Data

II.1 Context

This paper studies corruption in hiring in the context of a community health worker (CHW) program. Community health workers have become increasingly important in developing country health care systems, with over 5 million CHWs globally (Perry et al., 2014). In general, CHW programs attempt to extend health care services among under-served populations by providing targeted services for which a highly skilled provider is not required. In contrast to doctors and nurses, CHWs are usually hired without having an extensive background in health and are trained on the tasks they will perform. CHW programs often focus on promotion of health literacy, conducting basic check-ups, and distribution of nutrition supplements and medicines. CHWs may also take on more complicated roles, such as delivering babies, giving medical advice, or serving as tuberculosis treatment providers (DOTS).

The data in this study are from a rural area of a large developing country democracy. The study area contains slightly less than 2 million individuals and is served by around a thousand active CHWs. I study the hiring of approximately seventy supervisors, each of whom oversees 15-25 CHWs. Due to concerns about respondents and individuals who assisted the data collection being identified and retaliated against, the identity of the country was removed from the paper.

In the study area, all of the CHWs are women hired from the communities that they serve and trained by the government. The typical CHW is between 30 and 40 years of age, married (94%), has between 8 and 12 years of education, and has worked as a health worker for eight years. Depending on the local geography, a CHW typically provides services to between 850 and 1500 individuals. On average, CHWs earn slightly more than the median household income in the area and have substantial attachment to their work: sixty percent expect to remain in their current CHW position for the rest of their lives, while an additional 25% expect to remain as health workers, but eventually be promoted. This attachment is a function of their relatively high pay and near complete job security: in my data, only 1.1% had exited from the job either voluntarily or involuntarily over the course of a year.

The primary responsibility of these health workers is services for pregnant women and children. The average CHW serves 11.5 pregnant or recently delivered women at any one time, and visits them regularly. During the visits, they distribute iron supplements, provide basic antenatal care counseling, and perform post-natal checkups on newborns. They often serve as front-line health workers, with over half receiving visits more than once per week for medical advice. In a month, the median CHW distributes oral rehydration salt packages to six households and paracetamol tablets (used for reducing fever) to five. Slightly less than half of CHWs have served as tuberculosis treatment providers. [303x42]
treatment (DOTS) providers, and at the time of the survey, 25% had actively worked as a DOTS provider in the previous six months. While the CHWs provide additional services, these are some of their most important functions.

The studied hiring process was part of a wave of hiring of health worker supervisors across the country.\footnote{The study area was selected due to pre-existing relationships that enabled data collection, but the corruption in hiring seems to have been ubiquitous. While piloting the survey instrument in other regions, I encountered similar bribe-paying arrangements. More broadly, the presence of corruption in hiring is common knowledge across the country.} Prior to this hiring, CHWs lacked supervisors specifically tasked with monitoring their work, and national policymakers felt that more monitoring could improve training and effort provision. Only current CHWs were eligible to apply to become a supervisor, and so by collecting information on all the CHWs, I observe the universe of potential applicants. All current CHWs were grouped into geographically based clusters of 15 to 25 CHWs, and one supervisor was hired to oversee the work of each cluster. Crucially, only current health workers within a given cluster were eligible to apply for its supervisor position. All current CHWs were informed about the nature of the position and pay, and 34% of CHWs applied, with an average of 5.3 applicants per position.

Higher-level government officials attempted to standardize selection via a system in which applicants would be assigned points based on their education, past work as a CHW (e.g. assisting more than a certain number of institutional deliveries in the last six months), and an interview with the hiring committee. The applicant within a cluster who had the most points was supposed to be assigned the position, but as shown in section II.4, selection appears to have been based on bribery. The study area was composed of eight geographically contiguous regions with separate hiring committees composed of local health bureaucrats in each region. Each committee included the lead bureaucrat overseeing health programs in the region, as well as more junior bureaucrats. Given the lack of oversight and data at higher levels, it was easy to claim that the hire had the most points regardless of the truth.

After the supervisors were hired, there was a six month lag before their training due to administrative delays in preparing the training curriculum (see figure A.1 for a timeline). During this period, those selected as supervisors did not receive any training or take on any supervisory duties, but continued in the same role as CHWs. The eventual training focused on the tasks that they would carry out as supervisors, including their responsibilities when visiting CHWs in their villages, how to restock CHW supplies, strategies for “supportive supervision”, and how to conduct monthly meetings with CHWs. After the training, they began working as supervisors.

The duties of supervisors are: 1) go at least once a month to the village of each CHW to confirm that they are properly carrying out their work; 2) observe CHWs during their duties to provide feedback; 3) give feedback, based on centrally-defined criteria, on whether a given CHW’s performance is satisfactory; and 4) hold meetings at which they provide training and feedback. Based on data collected from the CHWs, supervisors typically visit CHWs in their villages between once and twice a month, with each visit lasting between 1-2 hours. During these visits, they almost always review the records and recent work of the CHW (96% of visits) and give feedback
and advice (81% of visits). In around half of visits, they directly observe the CHW’s work in the village, such as their counseling of pregnant women. CHWs generally appreciate their supervisors, with 71% stating that their supervisors are helpful or very helpful. Supervisors cannot fire CHWs and have little other sanctioning power, so their power to motivate workers is based primarily on verbal reprimands and informal incentives. Although perhaps less powerful than formal incentives, increased monitoring of government workers without formal incentivization has been found to be effective across numerous developing country contexts (Muralidharan et al., 2017a; Callen et al., 2018; Dal Bó et al., 2018).

Supervisors are compensated on a salary basis and earn around 40% more than they had as a CHW. This is the primary financial incentive to become a supervisor, as there are few opportunities to extract corrupt rents. Supervisors do not currently control any large flows of government funds from which they could skim and have little sanctioning power over their CHWs that they could use to extract bribes. As with CHWs, supervisors are rarely punished and job security is almost complete. 95% of supervisors stated they expect to remain as health workers for the rest of their lives, either in this position or at a higher level due to promotions, and after two years in the job, none had left or been fired. From discussions with current supervisors, their motivation to perform well comes from a combination of intrinsic motivation, fear of verbal reprimands from those who oversee the supervisors, and a hope that better performance will allow them to achieve even better jobs in the health bureaucracy.

II.2 Data collection

The paper relies on two main sources of data: two rounds of survey data collected by the author and administrative data from the government on CHW performance. Figure A.1 provides a project timeline, where data collection efforts are denoted in yellow and other events are in orange. The first round of survey data was collected after supervisors had been hired, but before they had started their new duties, while the second round was conducted six months after supervisors began their work. The first round focused on the work of CHWs over the preceding six months and hiring of supervisors. The second round was similar, but included questions about the performance of supervisors. During the surveys, I administered tests of health knowledge and general ability (Raven’s Progressive Matrices and a digit span memory test), psychometric instruments, a behavioral game measuring pro-social preferences, and a behavioral game measuring honesty. For both rounds of surveying, CHWs and supervisors were contacted via phone and made appointments to take the survey at a convenient central location. Respondents were compensated at 1.5 times the prevailing daily wage for participation and could earn more based on performance in the behavioral games, but were not told that prior to arrival. All supervisors and nearly all CHWs in the study area were surveyed (see appendix B for details).

I also use twenty months of monthly CHW-level administrative data on delivery of important health services. In each month, I observe the delivery of ten health services by each CHW, such as number of institutional deliveries assisted, number of newborn checkups conducted, and whether
they served as a tuberculosis DOTS treatment provider. The government aggregates these outcomes into a single performance measure between 0 and 100 that I focus on in the paper, but I also report results for individual services for robustness. Given the setting, one concern is possible manipulation of the data by supervisors. As a check, I compare the administrative data to performance evaluations that I collected for each supervisor from individuals overseeing the CHW program, but who were not involved in the selection of supervisors. Appendix figure A.4 shows that these measures, which could not be manipulated, line up with the administrative data (see appendix B.2 for details and additional validations).

Finally, I supplement these data with two auxiliary data sets (denoted as “Auxiliary Data” in figure A.1). The first consists of tests of reading ability and health knowledge that were administered to CHWs approximately a year before the hiring of supervisors, while the second is two years of monthly-level data on CHW salaries beginning approximately a year and a half prior to the hiring of supervisors. These data sets were collected in two regions containing around a quarter of CHWs in my sample and are used in robustness checks in appendix C.2.

II.3 Bribery data

The first object of interest is if hires paid bribes to get the job. Practically all papers on bribery rely on self-reporting (see Olken and Pande (2013) for a review), but respondents may not want to admit to illegal activity. To deal with potential underreporting, I combine information from multiple sources to determine payments. First, hires were asked if they had to pay any money for the job, and if so, how much they had given. 47% of hires told surveyors how much they had paid, while 33% declined to answer and 18% claimed that they did not pay anything. Given that a refusal to answer likely indicates the respondent paid a bribe, this suggests that at least 80% of hires paid to receive their position.

As an auxiliary source of data, I collected data from the other CHWs. Health workers and supervisors regularly socialize at trainings and health facilities, and after the hiring process, bribe payments were a popular topic of conversation. As a result, many CHWs had talked directly with their supervisors about the bribes that the supervisor had to pay. During the survey, the other CHWs were asked if they knew whether their supervisor had paid money for the job, how much their supervisor had paid (if anything), and where they had heard this information (e.g. supervisor, member of hiring committee). 77% of respondents stated that their supervisor had paid money, 19% said that they did not know, and just 3.8% claimed that their supervisor had not paid a bribe. Combining this with the direct reports from the supervisors, there is at least one bribe report per supervisor. For 97% of supervisors, either the supervisor herself admitted to paying a...

8The exact question wordings were “Many CHW supervisors told us that it was necessary to give some money to become a supervisor. Did you have to give anything?” and “How much did you have to give?”.

9In order to avoid making respondents uneasy, surveyors were instructed to not probe supervisors on this question: if the respondent did not give a response on the first query, they moved to the next question.

10The data suggest that the last group is wrong: in half of cases, the supervisor admitted to paying a bribe, and no supervisor had more than one CHW reporting that she did not pay a bribe.
bribe, or there are at least three independent reports of the supervisor’s bribe payment. Based on this, I conclude that all supervisors had to pay a bribe in order to get their job.

The second object of interest is the size of the payments. In cases where the supervisor was unwilling to tell the surveyor about her bribe payment, I estimate it based on the reports of other CHWs. 70% of CHWs who said their supervisor had paid a bribe also stated that they knew how much their supervisor had paid. If one or more respondents heard about the bribe directly from their supervisor, I privilege those direct accounts and take their average as the estimated bribe for that supervisor. In the remaining cases, I take the average of all respondents who said they heard their information from a reliable source such as a member of the hiring committee. In 92% of cases where the supervisor did not tell surveyors a bribe payment, secondary sources enable estimation; in most of these cases, the secondary sources heard directly from the supervisor. As a validation, I take cases in which the supervisor reported their payment and regress this on the bribe amount estimated from the secondary sources. As seen in the bottom panel of figure I, the fit is remarkably tight, with an $R^2$ value of 0.69, a slope of 0.83 ($se = 0.11$), and little systematic bias in the errors.

II.4 Understanding the market for jobs

Due to the difficulty of collecting accurate data on illegal activities, there is only a small empirical literature on contemporary markets for jobs. Two of the only examples are Wade (1982) and Wade (1985), which qualitatively describe how Indian irrigation engineers pay bribes to get favorable postings. However, these papers lack the quantitative data necessary for a more rigorous characterization of the market. A number of papers have shown that political connections can affect public sector hiring decisions (e.g. Fafchamps and Labonne, 2017; Jia et al., 2015; Colonnelli et al., 2019), but it is not clear how these connections interact with other factors such as bribery in generating hiring decisions. In this section, I use data on bribes, connections, and other applicant characteristics to demonstrate the presence of corruption in my setting and how competing patronage and bribery considerations may interact. Although this is a particular case, it has a wider relevance in demonstrating how these markets work.

The top panel of figure I shows the distribution of bribe amounts paid by hires, where the average bribe equals 17 months of supervisor salary. There are minimal opportunities for rent extraction in this context, and so hires are paying for the salary increase, career advancement, and non-pecuniary benefits of the job; the roughly 40% salary increase compensates for the bribe after a few years at local interest rates. The lack of opportunities for rent extraction may explain these payments are smaller than for jobs with more such opportunities (see footnote 2).

At a high level, corrupt hiring systems fall into three broad categories: (1) fully meritocratic, in which hiring decisions are made on the basis of merit, but then the hiring committee demands a

---

11 These reports are highly consistent: even for cases where the supervisor did not admit to paying a bribe, the correlation across reported bribe amounts for the same supervisor is 0.52.

12 Asking about the source was critical for accurate information. While CHWs with reliable sources had little systematic bias, those who said they had “heard it around” overestimated the bribe payment by a factor of two.
bribe from their preferred applicant; (2) fully non-meritocratic, in which hiring decisions are made solely based on bribes and connections; and (3) partially meritocratic, in which hiring decisions are made based on both meritocratic and non-meritocratic elements. To understand the relevant category for this process, I examine data on bribe offers from hires and those who were not hired. Although only those hired actually paid any money, unsuccessful applicants for the job were surveyed on whether they had been solicited for bribes and if they had made a bribe offer. 82% said that they had been solicited for a bribe offer, while 49% of unsuccessful hires told surveyors that they had made an offer for the job and how much they offered.

Applicants for the job differ in their bribe offers, political connections, and other characteristics. The hiring agent can be thought of as solving a discrete choice problem for each of the supervisor positions in which they select the applicant who maximizes their utility based on bribe offer, connections, and other characteristics. In the case of a fully meritocratic system, they would only put weight on meritocratic characteristics, while under a fully non-meritocratic system, only bribes and political connections matter. Appendix C.1 uses the revealed preference of the hiring agent in each of these competitions to estimate weights placed on bribes, connections, and other applicant attributes in hiring decisions. This analysis indicates that the hiring system was partially meritocratic, where the hiring decision is based not only on applicant bribe offers and political connections, but also a meritocratic element, applicant education. However, bribe offers were the most important factor, where they alone explain around 80% of the hiring decisions.

A partially meritocratic model is also more consistent with other evidence from survey data. Hiring agents solicited bribe offers from a large fraction (82%) of applicants; under a fully meritocratic system, it would only be necessarily to solicit bribes from the high quality applicants being considered for the job. During the survey, I also asked CHWs who did not apply for the job why they did not apply. By far, the most common reasons were that the CHW felt they did not have enough money (28.1%) or did not have enough education (48.4%), reinforcing that these were factors in hiring decisions. Section IV will provide further evidence against fully meritocratic selection, where the quality of hires and service delivery outcomes are related to wealth in a way that would not be expected under a fully meritocratic system.

---

13 A type of corruption that does not neatly fall into one of those categories is a disqualification system, as is sometimes seen in procurement contracts (Sukhtankar and Vaishnav, 2015). In this system, the selection committee solicits bribes from each candidate, and if a candidate refuses to pay a bribe, then they would be disqualified from consideration. The selection committee then may make the selection decision based on the quality of the candidates or other factors. This is clearly not a disqualification contest since only those hired actually paid a bribe.

14 The exact question wording was nearly identical to that for supervisors: “Many CHW supervisors told us that it was necessary to give some money to become a supervisor. Did you offer to give anything to become a supervisor?” I lack secondary sources to validate this data, as I did not think it was likely that information on offers would have circulated in a way that payments did. However, a number of robustness checks, including the high degree to which bribes offers are predictive of selection, suggests that this data is high quality (see appendix C.1).

15 At least one non-applicant mentioned money or education as a reason that they did not apply in 91% and 97% of competitions respectively.
III The effect of corruption on quality of hires

Given the important role of bribes in hiring decisions, existing literature suggests that the quality of hires may be poor (Shleifer and Vishny, 1993; Shleifer, 2004; Muralidharan, 2015; Sukhtankar and Vaishnav, 2015). As a first pass at testing this concern, I examine how hires compare to others who might have been hired. An attractive feature of this context is that since only current CHWs are eligible to apply, I not only observe hires and unsuccessful applicants, but also those who chose not to apply. It would typically be impossible to identify those who choose not to apply, since there would be no way to know who had considered applying for a given job. For corrupt processes, it is important to take non-applicants into account, as high quality individuals may choose not to apply under a corrupt process but might have applied under merit-based processes; if that were the case, it would be misleading to only compare hires to those who applied.

Table I tests for differences between hires, unsuccessful applicants, and non-applicants on a range of variables plausibly related to quality. These are divided these into three categories – hard skills (e.g. education, test scores), soft skills (e.g. honesty, prosociality), and past performance as a CHW. There are two main patterns. First, applicants are of a higher quality than non-applicants across all three categories (joint \( p \)-value < 0.001 for hard skills and past performance; joint \( p \)-value = 0.098 for soft skills). Second, those hired are of a higher quality than unsuccessful applicants in both hard skills and past performance as a CHW (joint \( p \)-value < 0.001; joint \( p \)-value = 0.028), while the difference is not statistically significant for soft skills (joint \( p \)-value = 0.75). The presence of corruption does not appear to preclude high quality candidates from applying or being hired.\(^{16}\)

While these results are suggestive, the true effect of interest is the difference in job performance between those hired under a corrupt system and those hired under counterfactual, non-corrupt systems. Given the infeasibility of an experimental approach,\(^ {17}\) I follow a different strategy: identifying who would have been hired under these counterfactual systems, and comparing these counterfactual hires to actual hires. Section III.1 discusses how to compare actual and counterfactual hires, while section III.2 develops a method for determining the counterfactual hires.

III.1 Construction of Supervisor Performance Index

In order to compare the actual and counterfactual hires, I determine the supervisor characteristics that are related to performance as a supervisor and compare actual and counterfactual hires based on an index of these characteristics. I measure supervisor performance based on the delivery of health services by the CHWs they supervise, where the better supervisors are those who motivate the greatest improvements in service delivery outcomes. I favor this measure over process measures

---

\(^{16}\) Appendix C.2 finds similar patterns using data on quality from prior to the hiring of supervisors, as well as addressing concerns related to the timing of data collection.

\(^{17}\) Outside of lab experiments, which may miss important aspects of corrupt hiring, ethical considerations preclude randomly assigning some places to undergo a corrupt hiring process. A different approach would be to randomly assign a crackdown on corruption in hiring for certain positions or in certain areas. However, to credibly identify the effect of corrupt hiring, it would be necessary to publicize the crackdown prior to candidates deciding whether to apply. Such publicity would almost certainly spillover onto the control group and invalidate the experiment.
of supervisory performance (e.g. number of meetings held with subordinates) since service delivery is the most important outcome for policy.\(^\text{18}\)

I use monthly administrative data on the delivery of ten health services by each CHW (e.g. institutional deliveries assisted, tuberculosis patients treated). The first month of data is from after the supervisor had been in place for two months, and the panel covers the following twenty months. The government aggregates these ten services into a single “functionality score” between 0 (worst) and 100 (best), which I focus on as a comprehensive measure of performance. For each health worker CHW \(i\) under a supervisor \(j\), I calculate \(\Delta y_{ij}\), equal to the average monthly change of CHW \(i\) on the functionality score over the twenty months of data. A positive value of \(\Delta y_{ij}\) indicates that the CHW was improving, while a negative value indicates that their performance declined. I then run a differences regression, regressing \(\Delta y_{ij}\) on characteristics \(v\) of each supervisor \(j\) in equation 1 in order to determine which supervisor characteristics are related to changes in delivery of health services.\(^\text{19}\) \(\beta_v\) is the average monthly change in the functionality score as a function of supervisor characteristic \(v\): a positive value indicates that CHWs under a supervisor with a higher value of \(v\) improved over time relative to CHWs under a supervisor with a lower value of \(v\). For example, if \(\beta_{\text{education}}\) were equal to 0.05, this would imply that the functionality score increased by an average of 0.05 more functionality points per month for CHWs under a supervisor with 12 years of education relative to CHWs under a supervisor with 11 years of education.

\[
\Delta y_{ij} = \alpha + \sum_v \beta_v X_{vj} + \epsilon_{ij} \tag{1}
\]

Since there are a large number of variables \(v\) and relatively few supervisors, I use LASSO for variable selection (Belloni et al., 2014; see appendix table A.5 for the full list of tested variables). The variables selected by LASSO are: 1) score on the Raven’s Progressive Matrices, a test of problem-solving ability; 2) years of education; 3) scores on tests of writing and reading abilities administered by surveyors; and 4) extroversion, as measured by the Big Five Index personality index. I use the estimated coefficients \(\beta_v\) from a post-LASSO regression to capture the relative importance of each variable (appendix table A.1) (Belloni et al., 2014) and aggregate the five indicators into an index to compare actual and counterfactual hires. This index is denoted as the predicted supervisor performance index (SPI) for the remainder of the paper.

Figure II plots the cumulative density function of SPI for: (1) non-applicants; (2) unsuccessful applicants; (3) hires; and (4) the predicted first best case, i.e. if the candidate with the highest value of SPI eligible for each job were selected. The index has been normalized so that the mean

---

\(^{18}\)Quality of hires is an important measure of allocative efficiency, i.e. was the position allocated to the highest quality candidate, but a corrupt hiring system entails possibly inefficient transfers from hires to existing bureaucrats. I focus on allocative efficiency since section III.3 shows that hire quality is far more important for social welfare than such transfers.

\(^{19}\)Since the data begins after the supervisors were in place for two months, I am unable to run a differences-in-differences (DID) regression. The current differences regression approach misses any immediate changes as a result of the supervisor’s presence. For a subset of outcomes in which a DID regression is possible, appendix section C.2 tests for an immediate response. There is a small immediate shift in hours worked per week, but not for any measures of service delivery.
and standard deviation of SPI for hires are equal to zero and one respectively, and higher values of SPI indicate better predicted performance as a supervisor. There is a strong positive selection on both the application and hiring margins. The median SPI value of SPI for the non-applicants is \(-1.97\) and for unsuccessful applicants is \(-0.33\), where the SPI distribution of unsuccessful applicants first-order stochastically dominates the distribution of non-applicants (Kolmogorov-Smirnov \(p\)-value < 0.001). Despite the selection on bribes, hires are of a much higher quality than the average applicant, where the SPI distribution of hires first order stochastically dominates that of unsuccessful applicants (Kolmogorov-Smirnov \(p\)-value = 0.001). It is also interesting that there is not much mass in the upper tail of the SPI distribution for non-applicants. This indicates that the need to pay bribes does not deter high quality individuals from applying.

The distribution of SPI among actual hires is relatively close to the distribution of SPI among predicted first-best hires, where one of the top three candidates was selected in 54\% of cases. However, there are some important differences between the actual hires and first best case. Although the corrupt hiring system delivered fairly good outcomes on average, a candidate in the bottom half of applicants for their position was selected in a quarter of cases. This demonstrates the potential for corrupt hiring practices to have heterogeneous impacts, as will be discussed further in section IV.

To quantify the magnitude of the effect of having a higher SPI supervisor, appendix table A.2 examines the relationship between SPI and delivery of four services (tuberculosis treatment, assisting institutional deliveries, newborn care, and nutritional counseling of pregnant women and new mothers). The dependent variable is the average monthly change in each of the services over the 20 months of administrative data for each CHW. Appendix table A.2 regresses that on the SPI of their supervisor and finds an economically meaningful relationship. For example, these estimates imply that after 20 months, a health worker's probability of serving as a tuberculosis treatment provider is 3.5 percentage points higher if they had been transferred from their supervisor to a different supervisor whose SPI is one standard deviation higher. Given that the mean probability of serving as a treatment provider in a given month was 20\% in the first three months of the data, this is substantial.

The key challenge with attributing these improvements to the supervisor is the possibility of endogenous matching of supervisors to clusters of CHWs. In particular, supervisors who are more educated or have higher Raven’s scores (i.e. higher values of SPI) may oversee clusters of CHWs that were already on an upward performance trajectory, in which case the improvements are not attributable to the supervisor. Note that since this is a differences regressions, bias will result from differential trends rather than differences in levels. To test for differential trends, I use data from my baseline survey on the number of institutional deliveries assisted by each CHW in each of the six months prior to the supervisors beginning in their role. Appendix figure A.2 plots the coefficients from a regression of the number of deliveries assisted by CHWs in each month on supervisor SPI and finds no evidence of pre-trends, suggesting that improvements indeed reflect the supervisor.

It is also helpful to consider the extent to which SPI captures supervisor quality, as there may
be unobserved characteristics that are important for supervisor performance, but are not included in SPI. To do so, I take three alternative measures of quality of the hired supervisors and test whether they provide additional information on supervisor quality above and beyond SPI, as would be consistent with SPI missing some important dimension of supervisor quality. These alternative measures of the quality of hires are: (1) performance evaluations of the supervisors by superiors (see appendix section B.2 for details); (2) average CHW rating of their supervisors on a scale from 1 to 4 (collected during the second survey); and (3) an index of process measures of supervisor performance. Panel A of appendix table A.3 regresses the change in CHW functionality score after the supervisor is hired ($\Delta y_{ij}$) on each of these, and finds the performance evaluations and average CHW rating both predict improvements, while the process measures do not. However, the alternative measures no longer have predictive power after SPI is added to the regression (column 4). This suggests that even if there may be other aspects of supervisor quality aside from that captured by SPI, SPI successfully captures a large fraction of supervisor quality.

III.2 Counterfactual estimation

The observed, corrupt hiring process produced a relatively high quality set of hires, but other hiring systems might do even better. In this section, I compare the actual hires to those who would have been hired under three merit-based hiring systems: (1) a test of health knowledge, (2) a test of problem-solving ability (Raven’s Progressive Matrices); and (3) the hiring rules that were supposed to be used, based on past performance as a health worker. I use a simple model to determine the set of counterfactual applicants and hires, and then compare the characteristics of the counterfactual hires to actual hires using SPI.

The key challenge is that the set of applicants under the observed, corrupt system may differ from applicants under merit-based systems, such as if honest individuals do not apply under a corrupt system but would apply under a merit-based one. Since only current CHWs are eligible to apply, I observe the universe of those considering applying for the job, including those whose decision to apply may differ under a merit-based system. A candidate will apply for a job if the expected benefit of applying exceeds the cost. The expected benefit is equal to how much she values the job ($v_i$) multiplied by the probability of being hired if she applies. Under a merit-based system, no bribes are paid, and so the cost of applying, $C_i$, is the cost of attending an interview or taking a test. There will exist some cut-off probability $e_i$ at which candidate $i$ is indifferent between applying and not applying for the job ($e_i v_i = C_i$). The decision rule for applying is simple: a candidate will apply under a particular hiring system if her probability of being hired is greater than $e_i$.

Determining the set of hires is a four step process. First, I use survey data and past application behavior to put bounds around $e_i$ for each candidate $i$. Second, I estimate the probability of each candidate being hired under each of the counterfactual hiring regimes if they choose to apply, and

---

20 This uses the second survey of CHWs on the frequency of supervisor interactions and number of supervisory tasks completed by their supervisor over the past two months.
third, I combine their cut-off probability and estimated probability of being hired to determine whether they would apply. Fourth, I examine the set of applicants for each position and determine which would be hired.

To build intuition prior to detailing the methodology, consider an example of a candidate in the health knowledge test counterfactual. In the first step, I determine this candidate’s cut-off probability $e_i$ is bounded between 0.25 and 0.33. In the second step, since this candidate scored a 96% on the health knowledge test and there are 15 other CHWs in their cluster eligible to apply, their estimated probability of being hired if they apply as 0.71. This is higher than 0.33, so they will apply for the job. I follow the same procedure for all the individuals in their cluster and determine that three others would apply, with test scores of 83%, 85% and 91%. The original individual has the highest test score, so is the counterfactual hire.

Moving into the methodology, the first step uses survey data to place lower ($e_l$) and upper ($e_u$) bounds around $e_i$ for each candidate. Using bounds is more robust to errors, so produces more credible estimates than trying to point identify $e_i$. During the survey, all respondents were asked if they would apply for the job as supervisor under various probabilities of being selected. The bounds are taken from the point at which they are no longer willing to apply, e.g. if they are willing to apply at a probability of 0.33, but not at 0.25, then $e_i$ must lie between those values (so $e_l = 0.25$ and $e_u = 0.33$). For those who applied in the observed hiring process, I use bribe offers to apply to tighten the upper bound on $e_i$.

Second, I estimate the probability of being hired for each candidate under each counterfactual. Selection under each counterfactual is based on some merit-based “score” such as a test score, where candidate $i$ is hired if her score $\kappa_i$ is the highest among the pool of individuals who applied for this job. I assume that each candidate knows her own score $\kappa_i$, but not the exact scores of individuals against whom she will compete. Instead, she knows the total number of individuals eligible to apply for the position, the distribution of scores among that population, $f(\kappa)$, and that her competitors have randomly and independently drawn their scores from a distribution $f(\kappa)$. Candidate $i$’s probability of being hired is equal to the probability that all of the other candidates in her cluster either draw lower values of $\kappa$ than her or prefer not to apply for the job.

Application decisions will depend on a candidate’s expectations of application strategies of other candidates, which makes the estimation of this probability challenging. To avoid those complications, I make a simplifying assumption: when a candidate considers the application behavior of individuals who draw higher values of $\kappa$ than her, she assumes those individuals will apply as long

---

21 The question wording was: “Think back to the time when you were deciding to apply. Suppose that you knew the hiring process would occur without giving money or using connections. If you knew that $|X|$ other women in your area were applying to become a supervisor, and you thought that you each had an equal chance, would you have applied?” If a respondent were willing to apply when facing two other women with equal chances, then this is a 1 in 3 chance), but not three other women with equal chances (1 in 4 chance), this implies that $0.33 > e_i > 0.25$. The survey asked this for an ascending number of competitors.

22 The bribe amount is a lower bound on an individual’s valuation of the job. Candidates agreed to participate in the survey for two hours in exchange for $\frac{1}{30}$ of a month’s salary, putting an upper bound on their cost of an hour of time. Taking 20 hours as a conservative upper bound on the time cost of applying under alternative systems, this implies that an individual offering a bribe of $b$ will apply as long as their probability of selection is greater than $\frac{b}{30}$. 

15
as they “want” the job, i.e. their value of the job is higher than the cost of applying. Thus her probability of being hired is equal to the probability that no other candidate draws a higher $\kappa$ than her and wants the job. Under this assumption, the probability of being hired is a simple function of a candidate’s score $\kappa$ and their number of opponents. I estimate this probability via simulation from the population of observed CHWs, generating 50,000 simulated clusters for each possible number of opponents by taking random draws with replacement from the full set of CHWs. For each of these simulated clusters, the “hired” candidate is the one with the highest score among those who want the job (based on a survey question on if they would apply if guaranteed to get the job without paying money). The cumulative density function of the scores of simulated hires is the estimated probability that a candidate with that score from a cluster of that size gets hired. For example, a candidate with a health knowledge test score of 96% is estimated to have a probability of 71% of being hired in a cluster of 15 CHWs, but only 53% in a cluster of 25 CHWs. For each candidate, I estimate their probability of being selected under each counterfactual system given their score and the number of candidates eligible to apply for the same supervisor position.

The simplifying assumption means that candidates slightly underestimate their own probability of being selected, as some competitors with higher values of $\kappa$ would choose not to apply. In practice, this does not affect the main results because the object of interest is who is hired for the job. The only application decisions that matter for who is hired is that of people who are likely to be hired. By definition, these individuals have high probabilities of being hired. Thus anyone who has a higher score than them will have an even higher probability of being selected, and so is quite likely to apply if at all interested in the job. As a result, this simplifying assumption turns out to be a good approximation. Previous drafts of this paper developed a fully structural method that accounts for strategic expectations of other candidate’s behavior. Results were almost identical, indicating that this assumption is fairly innocuous.

Third, I determine the sets of potential applicants using the bounds and estimated probabilities of being hired. When considering their application decisions, candidates can be split into three sets. The first is definite applicants, whose probability of being hired is larger than the upper bound on their cutoff and so will definitely apply for the job. The second is definite non-applicants, whose probability of being hired is below the lower bound on their cut-off and so will definitely not apply. For example, an individual with an upper bound of 0.33 and lower bound of 0.25 would be a definite applicant if their probability of being hired was 0.4 and a definite non-applicant if their probability of being hired was 0.05. The third set is possible applicants, whose estimated probability of being hired falls within their bounds. If the same hypothetical applicant had a probability of 0.29, they would be a possible applicant, since this probability falls between the upper bound of 0.33 and lower bound of 0.25, so it is not clear whether they would apply.

Finally, based on these three groupings, I determine the set of potential hires for each position, and based on that, place bounds around the SPI of the counterfactual hire. In each cluster, the applicant with the highest score would be hired. To determine the set of potential hires, I first take the individual with the highest score from the set of definite applicants. If all of the possible
applicants have a lower score than that individual, then that individual is the hire. On the other hand, if some of the possible applicants have higher scores than that individual, it is possible that they would have applied and been hired. In that case, the set of potential hires would consist of the definite applicant with the highest score and any possible applicants with at least that score. The upper and lower bounds on SPI for that particular job will be the maximum and minimum values of SPI from that set of potential hires. These bounds end up being quite tight since the strongest candidates are typically willing to apply at low probabilities and so are definite applicants.

III.3 Counterfactual results

To assess the relative quality of actual and counterfactual hires, I use the supervisor performance index. Since the scale of SPI is not intuitive, I translate it into a measure that is easier to interpret: the percent of first-best SPI captured by a particular hire. For a given cluster, this is equal to the SPI of the hire minus the lowest SPI of a candidate in their cluster, divided by the difference between maximum and minimum SPI of potential candidates in that cluster. This equals 1 if the highest SPI candidate in the cluster is selected and 0 if the lowest SPI candidate is selected; the first-best is to select the highest SPI candidate in each cluster, capturing 100% of the first-best case. Consistent with the results of section III.1, the actual hiring system is at 90% of the first-best case, indicating that it may be difficult for counterfactual systems to outperform it.

The first counterfactual is a standardized test of health knowledge. Governments frequently use standardized testing to hire civil servants, but tests will only select good candidates to the extent that test scores are related to job performance. This test was administered during the survey, and consisted of 30 questions taken directly from the CHW training manuals. Since health knowledge is directly relevant to the work of a CHW, this is a prime candidate as an alternative merit-based selection method. Hires under the health knowledge test are at 83-83.6% of the first-best case. Figure III plots these bounds for the counterfactuals, along with bootstrapped 95% confidence intervals around both the lower and upper bounds. The top bar of the confidence intervals is the upper bound on the confidence interval for the upper bound estimate and the lower bar is the lower bound of the confidence interval for the lower bound. The upper bound is 6.4 percentage points lower than first-best SPI under the actual corrupt system, and equivalence with actual hires can be rejected (lower bound \( p\text{-value} = 0.007 \), upper bound \( p\text{-value} = 0.008 \)).

The second counterfactual is the Raven’s Progressive Matrices, a test of cognitive ability that has been used in government hiring processes in Mexico and India. It is a shortened version of the full Raven’s matrices instrument (see appendix B for details) and was administered during the

\footnote{p-values are generated using a cluster bootstrapped paired t-test of difference in means in order to account for SPI being a constructed variable (Efron and Tibshirani, 1994). For the actual hires and counterfactual hires, I generate demeaned versions of each of the variables in SPI by subtracting out the within-group mean from each group (i.e. mean within actual and within counterfactual hires) and adding the overall mean. I then redraw samples with replacement at the cluster level, re-estimate SPI in each sample using the new variables, and save the t-statistic from a paired t-test of the differences between actual and counterfactual hires in each sample on this re-estimated SPI. The p-value is based on a comparison of the test statistic from a paired t-test between the actual and counterfactual hires to this bootstrapped t-statistic distribution.}
surveys. Hires selected using the Raven’s matrices are weakly better than the actual hires, where the percent of first-best SPI captured is bounded between 92.1% and 94.8%. While equivalence of SPI of the actual hires and upper bound of Raven’s hires can be rejected ($p$-value = 0.002), I cannot reject equivalence with the lower bound ($p$-value = 0.28).

The third counterfactual examines who would have been hired under the intended government procedure discussed in section II.1, which is the closest to the true counterfactual. The government created a formula to assign points based on factors such as past performance as a CHW, and the applicant with the most points was supposed to be hired. Although it was not possible to access the points as allegedly tallied by the hiring committee, since almost all of the points are based on performance as a CHW, I can approximate 65% of the possible points using administrative data and survey responses; the remaining 35% of points are from an interview with the hiring committee. The percent of first-best SPI under the intended government method is bounded between 86.4% and 89.3%, slightly lower than that of actual hires. While I can reject equivalence of the actual hires and the lower bound of hires under the government method ($p$-value = 0.087), I cannot reject equivalence with the upper bound of government hires ($p$-value = 0.938): this indicates that outcomes would have been no better and perhaps slightly worse under the uncorrupted procedure.24,25

Despite the corruption in the hiring process, hires are better than counterfactual hires under a health knowledge test and weakly better than hires under the intended hiring system. On the other hand, the Raven’s matrices outperformed the other systems, indicating that it is possible to get closer to the first best. These differences across counterfactuals are meaningful for service delivery outcomes: applying coefficients from appendix table A.2 (which examines the relationship between SPI and service delivery outcomes), a back of the envelope calculation indicates that relative to hires under the health knowledge test, the actual hires led to approximately 8% more tuberculosis patients being cared for and 2.5% more institutional deliveries, among other outcomes.

This can also be expressed economically using the value of a statistical life, which allows a comparison of the cost of potentially inefficient bribe payments to the gains from service delivery improvements. I conservatively focus only on the estimated increase in institutional deliveries since there is more extensive research on how that translates into lives saved. Combining the estimated relationship between supervisor SPI and institutional deliveries (section III.1; appendix table A.2), a conservative estimate of the relationship between institutional deliveries and reduced neonatal mortality from a meta-review (Tura et al., 2013), and a conservative valuation of a statistical life ($100,000), having a supervisor with 0.1 SD higher value of SPI improves social welfare by a value of approximately 60 salary-months annually. The average differences in SPI between hires under the health knowledge test, Raven’s matrices, and intended hiring system are 0.40-0.44 SD, 0.14-0.35 SD, and 0.01-0.23 SD respectively, which translates into meaningful differences in social

---

24This counterfactual omitted the interview score, which may be informative in selecting good supervisors. As a robustness check, I rerun the “best case” scenario for this counterfactual with interviews taken into account, in which interview scores are imputed as the candidate’s SPI. The estimated percent of first-best SPI is still bounded between 88.9% to 90.6%, which does not change the overall conclusion.

25Note that although the average quality is similar between the intended and corrupt systems, only 34% of the corrupt hires are selected under the intended hiring procedure.
welfare. Furthermore, when considering the costs and benefits of corruption, the value of hiring better supervisors far exceeds the value of bribe payments, suggesting that allocation of positions is the more important criterion to examine in evaluating counterfactual systems.

Robustness Checks As a robustness check on whether this methodology accurately predicts application decisions, I directly asked respondents whether they would apply if selection were based on a health knowledge test (without any corruption). When these responses are used to determine the set of applicants and hires, the resulting set of hires are at 83.4% of first-best SPI. This estimate falls right within my above estimates (83.0-83.6%) for the health knowledge counterfactual, consistent with the method successfully predicting application behavior. Results are generally robust to many different assumptions. Intuitively, this is because only the entry decisions of candidates with reasonably high probabilities of being hired end up mattering, as they are the ones who end up hired. Predicting the application behavior of those individuals is straightforward since almost all of the best candidates either would never apply or, more commonly, place a high value on the job. A high valuation means that they would be willing to apply even if their probability of being selected were relatively low. Since the best candidates by definition have a high probability of being selected, their estimated probability of selection is typically well above their application probability cutoff and so they choose to apply if they want the job. Thus even if the estimates of cutoff probability bounds or probability of being hired were slightly off, this would not affect the results.\textsuperscript{26,27}

IV Understanding efficient corruption

A broad question in economic development is how corruption affects economic outcomes. Although most empirical papers have found that corruption has harmful effects, the theoretical literature has highlighted ways that corruption can lead to efficient outcomes. For example, when a government allocates a limited number of “slots” among applicants (e.g. licenses, procurement contracts, jobs), if the socially optimal beneficiary of the slot has the highest willingness to pay, then under a corrupt allocation system, they would offer the largest bribe and receive the good (Leff, 1964; Beck and Maher, 1986; Lien, 1986). On the other hand, even if the socially optimal beneficiary has the highest willingness to pay, they may not be able to bid the most due to credit constraints, leading to an inefficient allocation towards the wealthy (Esteban and Ray, 2006).

\textsuperscript{26}For example, if perceived probability of selection were 50% higher than I had estimated, then under the government system, hires would be at 87.0-89.3% of the first-best case, almost identical to the estimated range of 86.4-89.3%.

\textsuperscript{27}One final concern with using SPI is whether the estimated relationship between personal characteristics and performance generalizes to counterfactual hires. Note that if supervisors were randomly selected, then the relationship clearly would generalize; the key is whether there are interactions between how supervisors are selected and how characteristics that make up SPI are related to performance. Panel B of appendix table A.3 finds no evidence of interactions between selection criterion (e.g. bribe size) and supervisor characteristics. Counterfactual hires are also similar to actual hires (e.g. 63-84% applied for the job, 40-56% gave a bribe offer), so the predictions are fairly in-sample.
In this section, I test how a key set of economic parameters affect the extent of misallocation under corrupt allocations. These parameters can be characterized within a simple model in which applicants apply for jobs allocated by a corrupt selection agent. In this model, which is described visually in appendix figure A.3, applicants are characterized by their level of quality, their wealth ($w$), and their valuation of the job.\textsuperscript{28} There are two quality types – high quality ($H$) and low quality ($L$) – with measure 1 of each type. The wealth of low quality applicants is distributed uniformly between $[0, 1]$, while that of high quality applicants is distributed uniformly over $[w, 1+w]$, where $-1 \leq w \leq 1$. For simplicity, a potential applicant’s valuation of the job takes on two values, high ($h$) and low ($l$), where $h > 1+w$ and $l < \min(0, w)$; in other words, high valuation bidders are credit constrained, while low valuation bidders do not want the job. I also assume the set of applicants with high valuations is measure 1, where a fraction $p$ of high valuation applicants are high quality and $(1 - p)$ are low quality. To highlight the relative importance of each factor separately, I will assume that valuation and wealth are independent, although this is easily relaxed.

Suppose that the number of jobs to be assigned is measure $S << 1$, and the agent uses a multi-unit, second-price auction\textsuperscript{29} to assign jobs to the highest bidders. High valuation candidates are credit constrained and cannot place bids greater than their wealth, meaning that their bids are equal to the maximum of 0 or their wealth level. Candidates with low valuations can be ignored since they do bid.

Under values of $w$, $S$ and $p$ such that at least some low quality and some high quality candidates are selected, the proportion of hires who are high quality is equal to $p + \frac{w}{S}(1 - p)$.\textsuperscript{30} The two parameters determining the relatively efficiency of outcomes are $w$ and $p$. Both $w$ and $p$ map directly to parameters that can be calculated in the data: the correlations of quality with wealth and valuation respectively.\textsuperscript{31} Since $p$ is bounded between 0 and 1, the model prediction is that the efficiency of the allocation increases as each of those correlations increases.

This model suggests that the effect of corruption is heterogeneous: corruption leads to relatively efficient outcomes in situations where applicant wealth and valuation are positively correlated with their quality (panel a of appendix figure A.3), but is relatively inefficient in cases where these are poorly aligned (panel b of appendix figure A.3). The institutional context will determine the nature of these correlations: for example, wealthier individuals may have higher levels of human capital that improve productivity (positive correlation), but might also be systematically biased against poorer households in delivery of services (negative correlation). The next section applies data from this context to test the hypothesis that the correlations of wealth, valuation, and quality determine

\textsuperscript{28}This is similar to the model in Banerjee et al. (2013), but focuses on testable implications in this context.

\textsuperscript{29}This simplifies the solution, but the allocation outcome will be the same under other standard types of winner-pay auctions.

\textsuperscript{30}In equilibrium, $S$ is equal to the sum of the high valuation and low quality candidates who are selected. Note that if $x$ is the lowest wealth of a candidate who is selected, then $S = \int_x^{1+w} p \, dy + \int_x^1 (1 - p) \, dy$. Solving for $x$, the lowest wealth candidates selected for the job will have wealth equal to $(1 - S + p w) / (1 - p)$. The fraction of hirings will be high quality and $(1 - p)(S - pw)$ will be low quality.

\textsuperscript{31}Specifically, the correlation of wealth and quality is equal to $w / \left( \sqrt{1 + w^2} \right)$ and the correlation of valuation and quality is $(p - 0.5)$. 


when corruption is allocatively efficient.

IV.1 Testing the hypothesis

To test whether these parameters determine the efficiency of outcomes, I take advantage of the externally determined, independent pools of candidates for each supervisor position: only current CHWs can apply to be supervisors, and they are restricted to applying to the one supervisor position that their work area falls under. This produces variation across supervisor positions in the correlations of wealth, valuation of the position, and quality among the set of potential applicants; in some pools of potential applicants, the best candidates tend to be wealthier than the average for that pool, whereas in others, they happen to be poorer. I measure these correlations, and test the model prediction that hires and service delivery outcomes are closer to the first-best case in clusters where the correlations are more positive.

I focus on the correlation of quality with wealth due to a lack of sufficiently granular data on valuation of the job.\textsuperscript{32} The wealth-quality relationship is empirically important, where one of the three wealthiest candidates is selected in 76\% of cases. Wealth is measured with survey data on assets and earnings,\textsuperscript{33} while the quality of a potential hire is measured using SPI. To remove any mechanical relationship between SPI and service delivery due to how SPI is constructed, I use “leave-out SPI” rather than SPI in calculating the correlation of wealth and quality. For each cluster, I estimate the relationship between supervisor characteristics and performance using all of the other clusters, create an index of these characteristics, and then use this index to calculate the wealth-quality correlation in the cluster. More precisely, for each cluster \( j \), I estimate an index \( SPI_{-j} \) using only data from other clusters, and calculate the correlation of wealth and \( SPI_{-j} \) in cluster \( j \). This correlation is among the set of individuals who would ever consider applying for the job rather than among those who applied since the application decision is endogenous.\textsuperscript{34} The correlation of wealth and quality varies substantially; it is negative in 26\% of clusters, and positive and greater than +0.2 in 64\% of clusters. This is consistent with heterogeneity in the quality of supervisors, where, for example, 26\% of hires were in the bottom half of applicants in quality.

Intuitively, this correlation is plausibly exogenous with respect to service delivery outcomes prior to the arrival of the supervisor. This is the correlation of wealth and SPI across CHWs in the cluster rather than the level of either variable, and so as shown below, clusters with a more positive wealth-SPI correlation are not wealthier, do not contain CHWs with higher average values.

\textsuperscript{32} Valuation is measured with survey questions on whether the CHW would apply for the job of supervisor under different probabilities of being selected, where those willing to apply under lower probabilities have higher valuations (see section III.2). This measure is too coarse and poor at differentiating individuals with high valuations to be used: nearly half of CHWs say they would apply at the lowest listed probability of selection.

\textsuperscript{33} I construct an index of CHW wealth using the first principal component of an index of household assets, earnings of the CHW and the earnings of other household members. There is substantial variation, primarily due to the earnings of their husbands: I find a sixfold difference between CHWs at the 10\textsuperscript{th} and 90\textsuperscript{th} percentiles of total monthly household earnings.

\textsuperscript{34} The correlation among those who do not want the job is irrelevant since they would never apply or be selected. Being interested in the job is defined as being willing to apply if they had a probability of 1 of getting the job upon applying and did not have to pay a bribe to get it.
of SPI, and have no better service delivery outcomes prior to the arrival of supervisors. There is no reason why this correlation across CHWs would have affected CHW productivity prior to the hiring of supervisors since CHWs work independently of one another. The correlation becomes relevant during the corrupt selection of supervisors, but would not have mattered previously since the supervisor position did not exist prior to this hiring.

The first test is whether candidate pools with a higher correlation of wealth and “leave-out SPI” across CHWs produce higher quality supervisors. Column (1) of Table II finds a strong and statistically significant positive relationship between quality of the hired supervisor and this correlation \( (p = 0.024) \). Translated into the units from the counterfactual section, these estimates imply that going from a candidate pool at the 10\(^{th}\) percentile of wealth-SPI correlation (a correlation of \(-0.26\)) to the 90\(^{th}\) percentile (a correlation of 0.66) is associated with moving from a hire who captures 84.4\% of the first-best SPI to one who captures 95.1\% of the potential first-best SPI.

As a second test, columns (2)-(6) of Table II evaluate whether candidate pools with a higher correlation of wealth and SPI experience larger increases in service delivery outcomes after the supervisor was hired. As seen in equation 2, the specification is a differences regression nearly identical to the one used in appendix table A.2. The dependent variable is the average monthly change in a service delivery outcome for CHW \( i \) in candidate pool \( j \) over the 20 months of administrative data. CHWs in pools with a stronger correlation of wealth and SPI experience statistically significant improvements in the index of services used by the government to measure overall CHW performance (“Functionality Score”). This can also be seen when broken down into individual services such as assisting with institutional deliveries, newborn check-ups, and nutritional counseling. To account for multiple hypothesis testing, I conduct a test of joint significance across the five service delivery outcomes (joint \( p \)-value=0.028). These relationships are also depicted visually in figure IV.

\[
\Delta y_{ij} = \alpha + \beta_{corr\_quality\_wealth_j} + \epsilon_{ij}
\]

The key assumption for these tests is that the wealth-SPI correlation causes these improvements through the identity of the selected supervisor rather than pre-existing characteristics of the cluster. In particular, there should not be pre-existing differences in service delivery trends that are related to the correlation of wealth and quality in a cluster since these are differences regressions (e.g. clusters with positive correlations were already on an upwards trend relative to those with negative correlations). Table III supports that the correlation of wealth-SPI in a cluster is plausibly

---

35 Prior to the hiring of supervisors, CHWs were administratively grouped into 3 to 6 CHWs that restocked their supplies at a common location. CHWs otherwise do not work together, so there is no possibility of complementarities across CHWs with respect to service delivery. Each supervisor oversees 20 CHWs formed from joining three to five geographically-proximate pre-existing groups.

36 There is significant positive selection into applying for the job, as seen in figure II. As a result, even when the correlation of wealth and SPI is negative, hiring outcomes are still reasonably good since most of the worst quality candidates did not apply.
exogenous, finding no evidence of a relationship between it and characteristics of CHWs in that cluster (joint \( p\text{-value}=0.57 \)). These potential confounders include characteristics of CHWs in the cluster (e.g. wealth, SPI) and baseline levels of service delivery. I measure baseline levels of service delivery with both data from the survey of CHWs prior to supervisors starting in the position, as well as the first month of service delivery outcomes reported in the administrative data (when supervisors have only been on the job for a short time, so potentially have not yet had time to influence CHW performance). I also test for pre-existing trends in service delivery outcomes as a direct test of the identifying assumption; it would be concerning if clusters with a stronger correlation of wealth and SPI were already on a positive trajectory prior to the hiring of supervisors. Using CHW-level data on institutional deliveries over each of the six months prior to the start of the supervisors, figure A.2 finds no evidence of differential pre-trends related to the correlation of wealth and SPI.\(^{37}\)

A separate concern is whether these relationships are attributable to the correlation of SPI with wealth rather than another variable that is correlated with wealth (e.g. Raven’s score). Note that the wealth-SPI correlation only becomes relevant for service delivery after the hiring of supervisors. Unlike wealth, which is important in corrupt hiring decisions, there is no reason for the correlation of another variable and SPI to affect service delivery post-hiring but not pre-hiring.\(^{38}\) This timing reinforces that the relationships are indeed attributable to the wealth-SPI relationship.

Given these pieces of evidence, I conclude that the service delivery improvements are attributable to the selection of better supervisors in clusters with a higher wealth-SPI correlation, in support of the mechanisms in the model. The high average quality of hires found in the previous section can be explained by the strong positive correlations between SPI and valuation and SPI and wealth across all health workers (+0.47 and +0.28 respectively). However, in clusters where the correlations were negative, a lower quality set of individuals tends to be hired. Thus even though outcomes are relatively efficient on average, the effects of corruption on allocational efficiency are heterogeneous.\(^{39}\) Despite this heterogeneity, the average effect calculated in the previous section is still meaningful: it can thought of as what would be expected for the many jobs like this one in which wealth, valuation, and quality are likely to be positively correlated.

\(^{37}\)One possible issue is if there were complementarities between a CHW’s wealth and value of SPI that affects their performance as a CHW (e.g. individuals with higher values of SPI are more personally effective as CHWs if they are wealthy), and so clusters with a higher correlation of those characteristics benefit from those complementarities. Appendix table A.4 regresses service delivery outcomes on CHW wealth, CHW SPI and the interaction of wealth and SPI. The interaction of wealth and SPI of individual CHWs is not predictive of service delivery outcomes, so this cannot explain the results.

\(^{38}\)The exceptions to this are education and political connections. To check this, I re-run the main specifications including the correlations of SPI with those variables as a regressor, but find almost no change in the magnitude and significance of the coefficient on wealth-SPI correlation when they are included.

\(^{39}\)These results also emphasize that the hiring process was not fully meritocratic. Under a fully meritocratic hiring process, wealth is not relevant for hiring decisions. Thus the correlation of wealth and quality in a pool of candidates should not determine the quality of the hires unless candidate pools with a stronger correlation have better candidates, which table III refutes.
IV.2 Discussion

This is one of the first papers to empirically demonstrate efficient corruption aside from Sukhtankar (2015) and Dreher and Gassebner (2013), which have different underlying mechanisms. The results of this paper go beyond those papers and the set of papers demonstrating the efficiency losses from corruption by showing how corruption can have heterogeneous effects depending on characteristics of the market. In particular, the efficiency consequences of corruption are predictable based on the correlations of wealth and valuation with the characteristics of the socially optimal beneficiary. This need not have been true: for example, if high quality individuals are unwilling to engage in bribery, then low quality individuals will be selected regardless of the correlations of wealth, valuation, and quality.40

In this market, outcomes were relatively allocationally efficient because the average correlations of quality with wealth and valuation were strongly positive. However, there are other types of jobs for which these signs may differ. Consider supervisory positions for police officers in high corruption environments. The wealthiest applicants will be officers who acted most corruptly in their current positions, and those with the highest valuation of the job will be individuals who anticipate extracting the most corrupt rents. This generates a negative correlation between quality and both wealth and valuation, and as in the candidate pools here with negative correlations, outcomes will be likely relatively inefficient. There is a long list of factors that will determine the sign of these correlations including ability to extract corrupt rents, expected length of tenure in the job, extent to which payment is based on performance, costs of performing well in the job, non-pecuniary motivations (e.g. status seeking), and inter-generational correlations of wealth.

The extent to which each matters will vary by context. In this case, human capital and ability were important for performance and are highly correlated with wealth. From discussions with CHWs, non-pecuniary motivations such as hoping to advance in the health bureaucracy and public service motivation may have been important for differences across CHWs in their valuation of the job as supervisor. The positive consequences of corruption here will probably generalize better to jobs in which there are fewer opportunities to extract corrupt rents (e.g. teachers, other health workers), and where previous wealth accumulation is a positive signal of quality (e.g. previous earnings are based on past job performance). These lessons could also be applied to other types of items allocated by governments (e.g. licenses to export or extract resources, in-kind transfers), where the same correlations govern the allocative efficiency of corruption.

This analysis focused on the static allocational effects of corruption, i.e. whether the jobs were allocated to the optimal set of hires. There are a number of additional factors to consider when thinking about the overall effect of corruption in hiring. First, there is the possibility of dynamic effects of corruption in hiring: if an individual expects to be able to purchase a job rather than earn it, this could reduce human capital investment. In the setting I study, the creation of the position of supervisor was a surprise, and so dynamic effects are not possible. To the best of my knowledge, 40

40 There may also be informational asymmetries across applicants that disrupt this relationship, and if hiring systems are purely patronage-based, these correlations are not the parameters that will govern efficiency of allocations.
there is no empirical evidence on this effect in other settings, but there are reasons to think that it would be small. If public sector jobs are scarce, it would be risky to under-invest in capital formation as it is uncertain whether one will be able to acquire a job corruptly. Furthermore, if there are complementarities between meritocratic and non-meritocratic considerations (Jia et al., 2015), there are still strong incentives to invest in human capital.

Second, there may be costs of corruption aside from the quality of hires, such as leading to economically regressive allocations, decreasing confidence in institutions, or degrading social norms. While hiring based on bribes is certainly regressive, merit-based hiring systems also tend to be regressive; counterfactual hires based on standardized tests were nearly as wealthy as the hires based on bribery, likely due to the human capital requirements for performing well on tests. Furthermore, the degradation in social norms from this form of corruption is also true of other forms of corruption. Given the impossibility of ending all corruption in the short-run, it is better to address its most costly forms. The paper provides evidence for a framework to understand when corruption will have the worst allocative consequences. The marginal dollar allocated to eliminating corruption in other areas may have similar effects on norm degradation as allocating it to a crack-down on corrupt hiring, but with greater efficiency benefits.

V Conclusion

This paper examines the effect of corruption on allocational efficiency. I collect data from a government hiring process and document the presence of substantial corruption. Despite that, few high quality candidates were deterred from applying and mostly high quality individuals were hired. This occurs because bribery leads to selection on wealth and valuation of the job, both of which are positively correlated with quality on average. However, in areas where the correlation between wealth and quality was negative, corrupt hires are of a lower quality.

These results have broader implications for policy. First, it is notoriously difficult for organizations to identify talent. Here, bribes were a second-best solution for signaling quality, but they have the unfortunate consequence of diverting resources from hires to corrupt superiors. Alternative hiring mechanisms that similarly select on valuation or wealth, such as ordeal mechanisms, may be superior. However, the paper also points to the potential danger of such a mechanism: if valuation is negatively correlated with quality, it would be preferable not to screen on it with an ordeal mechanism. Furthermore, these results emphasize the importance of screening on characteristics that are correlated with job-specific performance rather than generalized ability. As seen in the counterfactual exercises, screening mechanisms that seem ex ante similar may produce meaningfully different outcomes. Collecting data on bureaucratic performance to determine appropriate screening for hiring is a potentially low cost way to improve bureaucratic efficiency.

Second, the paper demonstrates empirically that corruption should not be thought of as categorically negative, but rather as having heterogeneous effects. Given the limited capacity of policymakers to fight corruption, governments should target anti-corruption resources towards the
sectors in which it leads to the biggest distortions. This paper offers a way to identify those sectors: for example, in the case of markets for jobs, the biggest distortions are for jobs that have substantial ability to extract bribes. These results do not imply that corruption in hiring should be ignored, but that given the difficulty of eliminating corruption, it is better to first address its most costly forms. Given the many papers finding large distortionary effects of corruption in other sectors (e.g. Olken 2007; Ferraz et al. 2012), it is likely more efficient to allocate anti-corruption efforts towards those areas.

This paper demonstrates a number of general economic mechanisms involved in corruption. In this context, it was possible to identify two components into the allocational efficiency of corruption, but it would be useful to validate in other contexts and study other influences. Examining different methods for selection of civil servants, particularly in mid-level management positions, is a fruitful area for future research.
References


NewsRu, “Medvedev admitted that government posts are sold in Russia, and Zhironovsky named the price,” *Novosti Rossii [Russian News]*, July 2008.


_ , Antonia Atanassova, Marianne Bertrand, and Sendhil Mullainathan, “Targeting with


Stancu, Elena, “Mass exodus: Why corruption in Romania’s healthcare system is forcing its doctors to work abroad,” The Independent, January 2014.


Figure I: Bribe payments

(a) Winning Bribe Amounts

(b) Actual vs. Predicted Bribe

Notes: The top sub-figure is the distribution of bribe payments that were made by those hired, where the payment unit is months of salary in the job that they are hired for. Since not all candidates were willing to report their bribes to the survey team, some bribe payments in the top sub-figure were estimated using reports from other sources. The bottom sub-figure checks the accuracy of those secondary reports. In cases where the bribe is directly reported by the hire, the figure compares the direct reports (y-axis) to the estimated bribe for that hire from the secondary reports (x-axis). All points would fall on the superimposed 45-degree line if the secondary reports were exactly accurate, but the errors are small and only slightly biased.
Figure II: Selection on Predicted Supervisor Performance Index

Notes: This figure plots the cumulative densities of predicted supervisor performance index (SPI) for non-applicants, unsuccessful applicants, those hired, and the predicted optimal set of hires. Comparisons between the groups show how the application and hiring decisions select on quality. The left-most line (lowest values of SPI) is for individuals who did not apply for the job. The second left-most line is for the set of individuals who applied for the job but were not hired. The third line is SPI for the set of individuals who were hired in the job. The final line takes the individual with the highest value of SPI from each of the competitions, i.e. the “predicted optimal” set of hires. All differences between groups are statistically significant at the 1% level in a Kolomogorov-Smirnov test for equality of distributions.
Notes: This figure plots the percent of first-best SPI under four hiring systems. The percent of first-best SPI for a particular position is equal to the SPI of the hire under that system minus the lowest possible SPI of a candidate for that position (including those who did not apply), divided by the difference between maximum and minimum SPI of potential candidates. This will equal 1 if the highest SPI candidate is selected and 0 if the lowest SPI candidate is selected for a given competition. I plot the mean of the lower and upper bounds across all of the positions. The first line is the SPI of hires under the actual, corrupt hiring process. The second is for hires under a test of health knowledge. The third is for hires under a test of problem solving ability (the Raven’s matrices). The fourth is the SPI of hires under the non-corrupt process that was supposed to be used for hiring, based on past performance as a health worker. I also plot 95% confidence intervals around the lower and upper bounds in each case, using bootstrapping to construct the confidence intervals. The top bar of the confidence intervals is the upper bound on the 95% confidence interval for the upper bound estimate and the lower bar is the lower bound of the 95% confidence interval for the lower bound. The inference procedure is modeled after Efron and Tibshirani (1993)’s methodology for bootstrapped two sample tests of differences of means. The idea is to generate a bootstrapped distribution of t-statistics under the null hypothesis of no differences in means between the actual and counterfactual hires, and then compare the estimated t-statistic to that distribution for the purposes of inference. For both the set of actual hires and the set of counterfactual hires, I demean the variables used to generate the Supervisor Performance index by subtracting out the within-group mean of each variable and adding the overall mean on that variable between the groups. The result is that if SPI were generated using these new variables, it would be equal between actual and counterfactual hires. I then redraw samples with replacement, where the resampling is done at the hiring cluster level to account for correlated characteristics of CHWs within clusters. In each sample, I re-estimate the supervisor performance index, but using the demeaned variables. I then conduct a paired t-test of the difference between the actual and counterfactual winners on this supervisor performance index, and save the value of the t-statistic. Across the samples, this provides the distribution of the paired t-statistic under the null hypothesis of no difference between actual and counterfactual hires, but taking into account the additional uncertainty from use of an estimated regressor. Finally, I calculate the t-statistic for a paired t-test of differences in means between the actual and counterfactual winner. I compare that to the bootstrapped distribution of t-statistics under the null hypothesis for inference.
Figure IV: Wealth-Quality Correlations and Service Delivery Outcomes

Notes: This figure examines the relationship between service delivery outcomes and the correlation of SPI and wealth within a given cluster. Each panel plots the average monthly change in a measure of service delivery in the cluster after the hiring of the supervisor against the correlation of wealth and SPI within that cluster. The first panel is of the “functionality score”, which aggregates the other service delivery outcomes into a rating between 0 and 100 via a government formula.
Table I: Characteristics of non-applicants, unsuccessful applicants, and successful applicants

<table>
<thead>
<tr>
<th></th>
<th>Non-applicant mean</th>
<th>Unsuccessful applicant mean</th>
<th>Successful applicant mean</th>
<th>Col 1 vs col 2 p-value</th>
<th>Col 1 vs col 3 p-value</th>
<th>Col 2 vs col 3 p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Hard Skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health Knowledge</td>
<td>.46</td>
<td>.57</td>
<td>.63</td>
<td>0.001</td>
<td>0.001</td>
<td>0.082</td>
</tr>
<tr>
<td>Short Term Memory</td>
<td>4.3</td>
<td>4.6</td>
<td>4.9</td>
<td>0.000</td>
<td>0.000</td>
<td>0.016</td>
</tr>
<tr>
<td>Raven’s Score</td>
<td>2.5</td>
<td>3.7</td>
<td>4.6</td>
<td>0.000</td>
<td>0.000</td>
<td>0.185</td>
</tr>
<tr>
<td>Education</td>
<td>8.8</td>
<td>11</td>
<td>13</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Reading Skill</td>
<td>4.4</td>
<td>5.5</td>
<td>5.8</td>
<td>0.000</td>
<td>0.000</td>
<td>0.115</td>
</tr>
<tr>
<td>Writing Skill</td>
<td>4.1</td>
<td>5.3</td>
<td>5.8</td>
<td>0.000</td>
<td>0.000</td>
<td>0.029</td>
</tr>
<tr>
<td><strong>Joint p-value</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Panel B: Soft Skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro-social (Donation)</td>
<td>.23</td>
<td>.22</td>
<td>.24</td>
<td>0.289</td>
<td>0.504</td>
<td>0.419</td>
</tr>
<tr>
<td>Honesty (Dice Game)</td>
<td>.49</td>
<td>.46</td>
<td>.46</td>
<td>0.169</td>
<td>0.092</td>
<td>0.321</td>
</tr>
<tr>
<td>Public Service Motivation</td>
<td>3.8</td>
<td>3.9</td>
<td>3.9</td>
<td>0.092</td>
<td>0.063</td>
<td>0.155</td>
</tr>
<tr>
<td>Intrinsic Motivation</td>
<td>3.3</td>
<td>3.6</td>
<td>3.6</td>
<td>0.009</td>
<td>0.008</td>
<td>0.027</td>
</tr>
<tr>
<td>Extrinsic Motivation</td>
<td>2.9</td>
<td>2.9</td>
<td>2.9</td>
<td>0.012</td>
<td>0.004</td>
<td>0.885</td>
</tr>
<tr>
<td><strong>Joint p-value</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.098</td>
<td>0.058</td>
<td>0.752</td>
</tr>
<tr>
<td><strong>Panel C: Past Performance as a Community Health Worker</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOTS Patients</td>
<td>1.1</td>
<td>1.5</td>
<td>1.8</td>
<td>0.026</td>
<td>0.111</td>
<td>0.890</td>
</tr>
<tr>
<td>CHW Work Hours (per week)</td>
<td>14</td>
<td>16</td>
<td>19</td>
<td>0.008</td>
<td>0.002</td>
<td>0.162</td>
</tr>
<tr>
<td>Client Visit Tasks</td>
<td>3.8</td>
<td>4.1</td>
<td>4.3</td>
<td>0.002</td>
<td>0.008</td>
<td>0.410</td>
</tr>
<tr>
<td>Medical Advice Frequency</td>
<td>4.9</td>
<td>5.1</td>
<td>5.5</td>
<td>0.771</td>
<td>0.144</td>
<td>0.451</td>
</tr>
<tr>
<td>Contraception Work</td>
<td>2</td>
<td>2.4</td>
<td>2.9</td>
<td>0.169</td>
<td>0.063</td>
<td>0.892</td>
</tr>
<tr>
<td>Immunization Work</td>
<td>1.2</td>
<td>1.3</td>
<td>1.4</td>
<td>1.000</td>
<td>0.845</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>Joint p-value</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.028</td>
</tr>
</tbody>
</table>

**Notes:** This table compares hires, the median non-applicants and the median applicant from each competition on a variety of variables individually. It uses a Wilcoxon signed-rank test to sign selection across these two margins for a large set of variables that are plausibly correlated with performance in the job. The first three columns give the overall mean of the three groupings of candidates, while the last three give the p-value from a Wilcoxon signed rank test comparing the groups. Health knowledge refers to the percent of questions answered correctly on a test of health knowledge administered during the survey. Short-term memory is the maximum number of digits successfully recalled during a digit-span memory test. Raven’s score refers to the number of correct answers (out of 12) on the short-form Raven’s matrices test. Education is years of education. Reading and writing skills are scores out of 6 on a test of reading and writing ability. Pro-social (donation) is the fraction donated in a dictator game to an orphanage, while honesty is the number of dice rolls other than fives and sixes reported on a modified Hanna and Wang (2017) honesty task. Public service motivation, intrinsic motivation, and extrinsic motivation are on a five point scale, with five as the highest score. DOTS patients is the number of DOTS patients served by the CHW. CHW work hours are hours per week worked as a CHW in the baseline survey, while client visit tasks are the number of tasks carried out when visiting clients (e.g. providing information about pregnancies and conducting check-ups). Medical consultation frequency is how often they are consulted for medical advice, where higher values indicate more frequent consultations. Contraception and immunization work indicates the number of households assisted in the last week with those needs. Joint p-values are based on the joint test of Kling et al. (2007) across all of the variables in the category. * p < 0.10, ** p < 0.05, *** p < 0.01
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervisor SPI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functionality Score</td>
<td>.97**</td>
<td>.709**</td>
<td>.0221**</td>
<td>.0105*</td>
<td>.00866*</td>
<td>.000124</td>
</tr>
<tr>
<td>Institutional</td>
<td>(.43 )</td>
<td>(.33)</td>
<td>(.011)</td>
<td>(.0055)</td>
<td>(.0046)</td>
<td>(.0041)</td>
</tr>
<tr>
<td>Delivery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newborn Check-ups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nutritional Counseling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOTS provider</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table examines the relationship between the correlation of wealth and SPI in a cluster and later service delivery outcomes for that cluster. Column (1) tests whether this correlation is related to the quality of the selected supervisor, while the remaining four columns measure the CHW-level average monthly change in four service delivery outcomes in the administrative data. Since SPI is a constructed regressor, standard clustered standard errors will overestimate the true level of precision. To incorporate the additional uncertainty from construction of the index, I use cluster bootstrapped standard errors, where SPI is reconstructed in each bootstrap sample based on that sample of data. For the joint test, I use resampling to generate the distribution of the joint test statistic under the null hypothesis (permuting the value of wealth among CHWs in the same cluster on each random draw to eliminate any relationship between the correlation of wealth and SPI and performance changes). I again reconstruct SPI in each permutation sample to account for additional uncertainty from the construction of the index. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 

Observations 66 930 930 930 930 930
Joint $p$-value .028
Table III: Placebo test for cluster-level wealth-quality correlations

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wealth</td>
<td>-.0053</td>
<td>(.026)</td>
<td>[.84]</td>
</tr>
<tr>
<td>Supervisor Performance Index</td>
<td>.24</td>
<td>(.42)</td>
<td>[.56]</td>
</tr>
<tr>
<td>Raven’s Score</td>
<td>.048</td>
<td>(.26)</td>
<td>[.85]</td>
</tr>
<tr>
<td>Education</td>
<td>.25</td>
<td>(.35)</td>
<td>[.47]</td>
</tr>
<tr>
<td>Reading Skill</td>
<td>.13</td>
<td>(.24)</td>
<td>[.6]</td>
</tr>
<tr>
<td>Writing Skill</td>
<td>.034</td>
<td>(.27)</td>
<td>[.9]</td>
</tr>
<tr>
<td>Health Knowledge</td>
<td>-.00063</td>
<td>(.043)</td>
<td>[.99]</td>
</tr>
<tr>
<td>Government Hiring System Points</td>
<td>-.042</td>
<td>(.28)</td>
<td>[.88]</td>
</tr>
<tr>
<td>Monthly Deliveries (Baseline)</td>
<td>-.21</td>
<td>(.25)</td>
<td>[.4]</td>
</tr>
<tr>
<td>Initial Newborn Visits (Admin)</td>
<td>-.037</td>
<td>(.073)</td>
<td>[.61]</td>
</tr>
<tr>
<td>Initial Nutrition Counseling (Admin)</td>
<td>-.06</td>
<td>(.062)</td>
<td>[.34]</td>
</tr>
<tr>
<td>Initial Deliveries (Admin)</td>
<td>-.25</td>
<td>(.24)</td>
<td>[.31]</td>
</tr>
<tr>
<td>Initial DOTS Provider (Admin)</td>
<td>-.074</td>
<td>(.056)</td>
<td>[.19]</td>
</tr>
<tr>
<td>Initial Functionality Score (Admin)</td>
<td>-7.6</td>
<td>(5)</td>
<td>[.13]</td>
</tr>
<tr>
<td>Monthly Deliveries (Pre-trend)</td>
<td>-.014</td>
<td>(.096)</td>
<td>[.88]</td>
</tr>
</tbody>
</table>

Observations 986
Joint p-value .57

Each row gives the coefficient, standard error, and p-value from a regression of individual CHW characteristics on the correlation of wealth and quality within the cluster. "Government Hiring System Points" is the number of points that the CHW had under the intended government counterfactual hiring system. "Monthly Deliveries (Baseline)" comes from the baseline survey done with CHWs prior to the hiring of the supervisors, while the other measures of service delivery are the value of that outcome in first time that the CHW is observed in the government administrative data. Standard errors are in parentheses and clustered at the supervisor level, while p-values are in brackets. Joint p-value comes from an F-test of joint significance. * p < 0.10, ** p < 0.05, *** p < 0.01.
For Online Publication

A Online Supplemental Appendix Tables and Figures

Table A.1: Post-LASSO regression on CHW performance

<table>
<thead>
<tr>
<th>Functionality Score</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>0.11**</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
</tr>
<tr>
<td>Raven’s Score</td>
<td>0.063**</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
</tr>
<tr>
<td>Writing Skill</td>
<td>0.20**</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
</tr>
<tr>
<td>Reading Skill</td>
<td>0.43**</td>
</tr>
<tr>
<td></td>
<td>(0.202)</td>
</tr>
<tr>
<td>Extroversion</td>
<td>0.30***</td>
</tr>
<tr>
<td></td>
<td>(0.101)</td>
</tr>
</tbody>
</table>

Observations 917
Mean 0.058

Notes: This table takes supervisor characteristics selected by LASSO and regresses them on the average change in functionality score for the CHW over the 20 months of administrative data. Standard errors are clustered at the supervisor level. * p < 0.10, ** p < 0.05, *** p < 0.01
Table A.2: Supervisor performance index and health services delivery

<table>
<thead>
<tr>
<th></th>
<th>Institutional Delivery</th>
<th>Newborn Check-ups</th>
<th>Nutritional Counseling</th>
<th>DOTS provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted SPI</td>
<td>.0077** (.0035)</td>
<td>.0053*** (.0016)</td>
<td>.0059*** (.0015)</td>
<td>.0018* (.0011)</td>
</tr>
<tr>
<td>Observations</td>
<td>917</td>
<td>917</td>
<td>917</td>
<td>917</td>
</tr>
<tr>
<td>Initial Mean</td>
<td>2.598</td>
<td>0.784</td>
<td>0.109</td>
<td>0.200</td>
</tr>
<tr>
<td>Effect Size (1 SD)</td>
<td>0.155</td>
<td>0.107</td>
<td>0.119</td>
<td>0.0351</td>
</tr>
</tbody>
</table>

Notes: This table studies how performance changes over a 20 month time period under a supervisor with a higher value of the characteristics in the SPI index. Coefficients are equal to the average monthly improvement on the outcome of interest for a CHW whose supervisor has a one SD higher value of SPI. Initial Mean is equal to the mean value for delivery of that service across CHWs for the first three months of data. Effect Size (1 SD) translates the coefficient into the estimated effect on service delivery over the 20 months of data. Column (1) refers to the number of institutional deliveries that the community health worker assisted in a month. Column (2) is the fraction of newborn children upon whom the community health worker conducted a check-up in the month. Column (3) is the percentage of pregnant women in their catchment area that the community health worker visited and provided nutritional counseling in the month. Column (4) is a binary variable for whether the community health worker was serving as a tuberculosis treatment provider during this month. Since SPI is a constructed regressor, clustered standard errors will overestimate the true level of precision. To incorporate the additional uncertainty from construction of the index, I use cluster bootstrapped standard errors, where SPI is reconstructed in each bootstrap sample based on that sample of data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table A.3: Supervisor Performance Index Robustness Checks

<table>
<thead>
<tr>
<th>Functionality Score (1)</th>
<th>Functionality Score (2)</th>
<th>Functionality Score (3)</th>
<th>Functionality Score (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Comparison of Measures of Supervisor Quality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance Evaluation</td>
<td>0.117***</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.031)</td>
<td></td>
</tr>
<tr>
<td>CHW Rating</td>
<td>0.740*</td>
<td>0.383</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.387)</td>
<td>(0.253)</td>
<td></td>
</tr>
<tr>
<td>Process Rating</td>
<td></td>
<td>0.109</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.080)</td>
<td></td>
</tr>
<tr>
<td>Supervisor SPI</td>
<td></td>
<td>0.500***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.060)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>917</td>
<td>917</td>
<td>917</td>
</tr>
<tr>
<td>Panel B: Interactions Between Supervisor Performance Index and Selection Criterion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor SPI</td>
<td>0.542***</td>
<td>0.495**</td>
<td>0.511***</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.227)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>Supervisor SPI X Bribe</td>
<td>0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor SPI X Connection</td>
<td>0.109</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervisor SPI X Education</td>
<td>0.045</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>917</td>
<td>906</td>
<td>917</td>
</tr>
</tbody>
</table>

This table conducts robustness checks related to the use of the Supervisor Performance Index in the paper. Panel A examines the extent to which SPI captures supervisor quality, as there may be unobserved elements of supervisor quality that SPI does not measure. The first three columns of Panel A regress average monthly change in the functionality score of CHWs after the supervisor is hired on three different types of evaluations of the supervisor. In order, these evaluations are: (1) performance evaluations of the supervisors by individuals overseeing the program, (2) ratings of the supervisor by the CHWs they supervise, and (3) an index of process measures of supervisor performance (frequency of interactions with CHWs, supervisory tasks completed over the past two months). Column (4) of panel A tests whether the evaluations that are predictive of service delivery changes contain additional information about the supervisor after SPI is accounted for. Panel B studies whether there are interactions between the characteristics used to select supervisors and SPI in the production of improvements in CHW service delivery. This is used to test the validity of extrapolating SPI to non-hires by testing for interactions between how supervisors are selected and how SPI is related to performance. In all columns, the dependent variable is average monthly change in functionality score of the CHW after the supervisor is hired. Column (2) examines the interaction between the size of bribe paid by the supervisor and SPI, while column (3) examines the interaction with whether the supervisor used a connection to be hired. Column (4) tests for an interaction with supervisor education, but the strong collinearity with SPI renders the test uninformative. In both panels, standard errors are clustered at the supervisor level. * p < 0.10, ** p < 0.05, *** p < 0.01.
Table A.4: CHW Wealth, SPI and Changes in Service Delivery Outcomes

<table>
<thead>
<tr>
<th></th>
<th>(1) Functionality Score</th>
<th>(2) Institutional Delivery</th>
<th>(3) Newborn Check-ups</th>
<th>(4) Nutritional Counseling</th>
<th>(5) DOTS provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wealth X CHW SPI</td>
<td>0.0523 (0.0533)</td>
<td>-0.000165 (0.00326)</td>
<td>0.000271 (0.000897)</td>
<td>0.000692 (0.000832)</td>
<td>-0.000575 (0.000810)</td>
</tr>
<tr>
<td>Dependent mean</td>
<td>.079 (0.0533)</td>
<td>-.033 (0.00326)</td>
<td>-.0037 (0.000897)</td>
<td>-.0024 (0.000832)</td>
<td>-.0035 (0.000810)</td>
</tr>
<tr>
<td>Observations</td>
<td>882</td>
<td>882</td>
<td>882</td>
<td>882</td>
<td>882</td>
</tr>
<tr>
<td>Joint p-value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.33</td>
</tr>
</tbody>
</table>

Notes: This table regresses average monthly changes in service delivery outcomes after the supervisor was hired on CHW wealth, CHW SPI, and the interaction of CHW wealth and SPI. The table gives the coefficient for the interaction term. The joint p-value is based on an F-test of joint significance. Standard errors are clustered at the supervisor cluster level. * p < 0.10, ** p < 0.05, *** p < 0.01.
Figure A.1: Timeline of Project and Data Collection

Months from Supervisors Beginning Their Job

-30 -20 -15 -10 -5 0 5 10 15 20 25

- CHW Supervisor Hiring Announced
- Hiring Decisions Announced
- First Survey of CHWs/Supervisors
- Training of CHW Supervisors
- CHW Supervisors Begin Work
- Second Survey of CHWs/Supervisors
- CHW Service Delivery Admin Data
  - Auxiliary Data
    - CHW Health Knowledge Test
    - CHW Salary Data
Figure A.2: Test for Pre-Trends in Institutional Deliveries

(a) Pre-Trends related to SPI of the supervisor

(b) Pre-Trends related to the cluster-level correlation of wealth and SPI

Notes: This figure tests for pre-trends in service delivery outcomes that are related to the SPI of the hired supervisor (panel A) and the correlation of wealth and SPI of the hired supervisor (panel B). It plots the coefficients from a regression of the number of institutional deliveries assisted by a CHW in each of the six months prior to the start of the supervisor on the relevant regressor. The excluded month in this regression is the month prior to the start of the supervisors. Confidence intervals are at the 90% and 95% levels. Standard errors are clustered at the candidate pool level.
Notes: This figure illustrates the parameters of the model in section II.4. In the model, there is measure 1 of both low and high quality candidates applying for measure $S$ jobs. Candidate wealth is uniformly distributed between $[0, 1]$ for low quality candidates and $[w, 1+w]$ for high quality candidates, where $-1 < w < 1$. Candidates either have a high or low valuation for the job, where a proportion $p$ of high quality candidates and $(1 - p)$ of low quality candidates place a high valuation on the job. The left square represents the low quality candidates and the right square is the high quality candidates. The x-axis represents valuation (which I discretize into low and high for simplicity in the model), the y-axis represents wealth, and candidate valuation/wealth is uniformly distributed over the square. The low valuation candidates are unshaded, while the high valuation candidates are shaded. Those candidates who are hired are marked by the diagonal lines, i.e. those above a wealth of $1 - S + pw$.

In the top figure, there is a positive correlation between wealth and quality ($w > 0$) and a positive correlation between valuation and quality ($p > 0.5$), while in the bottom, there is a negative correlation between wealth and quality ($w < 0$) and a positive correlation between valuation and quality ($p > 0.5$). As the correlation between wealth and quality decreases, the proportion of hires who are high quality also decreases.
Figure A.4: Performance Evaluations and Administrative Data

Notes: This figure compares administrative data on the performance of CHWs supervised by a given supervisor against independent performance evaluations of those supervisors to check for manipulation of the administrative data. On the y-axis is the average monthly change in functionality score for their CHWs after the supervisor was hired over the 20 months of administrative data. The independent performance evaluations are on the x-axis and come from individuals overseeing the supervisors who were not involved in hiring. The performance evaluations were along nine dimensions: overall quality of supervisor, quality of meetings with CHWs, health knowledge, overall intelligence, selflessness, interest in improving health outcomes, competitiveness, and desire to help CHWs improve. Independent Performance Evaluation Rating is the first principal component of those evaluations.
B Supplemental Material on Data Collection

B.1 Survey Details

The author managed all survey operations. Respondents were given 150% of average CHW daily earnings and had the potential to earn up to an additional 75% based on their performance in behavioral games, but they were not told this prior to arrival. Since the survey took between 1-2 hours, and payment exceeded the typical daily wage rate, refusal rates were very low. All surveyors were female since the CHWs are female, and most had previous experience surveying this population of CHWs.

The first round of surveying occurred after supervisors had been hired, but before they had started their new duties. In this round, a random sample of CHWs and 98% of supervisors were interviewed. This survey focused on the work of CHWs over the preceding six months, and administered a test of health knowledge and numerous psychometric instruments. It also asked about the hiring of supervisors. The second round of surveying was six months after supervisors began their work and attempted to interview all CHWs and supervisors. This survey focused on the performance of supervisors and CHWs, as well as administering two tests of general ability (Raven’s Progressive Matrices; digit span memory test), a test of health knowledge, a behavioral game measuring pro-social preferences, and a behavioral game measuring honesty. We were able to contact 96.4% of the sample frame, and of those contacted, 92% were administered the survey.\footnote{Among those that we were not able to contact, some have likely discontinued their work as a CHW. However, in cases where we were unable to confirm this, I leave them in the sample frame. In cases where it was not possible to contact a particular CHW by phone, we attempted to contact them via other health workers who lived near them. Outright refusal rates were very low (0.5%), with most attrition due to being out of town or family obligations.}

The remainder of the section describes tests and behavioral games used during data collection.

Ability (Problem solving): The Raven’s Progressive Matrices measure general cognitive ability and have been used in hundreds of academic papers, as well as by some government agencies (e.g. Dal Bó et al. (2013); similar problems have been included in Mexican and Indian civil service exams). The test consists of a series of visual patterns of abstract shapes. From each pattern, a piece is missing, and respondents must identify the missing piece from a list of options. To induce effort, respondents were paid for each correctly answered problem and could earn up to a third of the prevailing daily wage. The set of 12 matrices were taken from the Advanced Progressive Matrices, Set I, as published by Pearson Clinical.

Short-term Memory: In the digit span memory test, surveyors recite a string of digits (e.g. 1-8-3-4-5) to the respondents and ask them to repeat it in same order. Following this, they are given a second string containing the same number of digits, and again have a single opportunity to give a correct answer. If either response is correct, then the number of digits increases by one, and the process repeats until the respondent cannot successfully repeat either opportunity for a given number of digits. The longest number of digits correctly repeated is their score.

Pro-social Preferences: Pro-social preferences were measured via a modified dictator game, which other studies have found is predictive of real-world pro-social actions (e.g. Lagarde and Blaauw, 2014). Respondents were informed that after the survey was completed, we would select sixteen respondents and give them an an amount approximately equal to a third of their average
monthly earnings. If they desired, they could donate some fraction of this to a local orphanage, but they had to decide this at the time of the survey, prior to finding out if they had won. Respondents were given this amount in fake local notes to split between envelopes marked “donation” and “self”.

Dishonesty: Honesty was measured using a modified version of a behavioral game from Hanna and Wang (2017). After completing the survey and other games, CHWs were given a dice and told to roll it 40 times, noting their rolls on a sheet. They were told that for each roll of 5 or 6, they would receive a payment, but for any other roll, they would not receive anything. They could earn slightly less than half the prevailing daily wage if they reported all 5’s or 6’s. Respondents have an incentive to act dishonestly by reporting a larger number of 5’s and 6’s, and can do so without fear of repercussion, since it is never possible for others to know conclusively whether they cheated. This game was played after the survey to avoid priming on dishonesty. 58% reported winning rolls above the 95th percentile of what would be expected by chance.

Table A.5: Variables included in LASSO for creation of SPI

<table>
<thead>
<tr>
<th>Category (# unique variables)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Knowledge (1)</td>
<td>Score on a 30 question test of health knowledge</td>
</tr>
<tr>
<td>Digit span Test (1)</td>
<td>Score on digit span memory test</td>
</tr>
<tr>
<td>Raven’s Matrices (1)</td>
<td>Score on Raven’s Progressive Matrices</td>
</tr>
<tr>
<td>Education (1)</td>
<td>Years of Education</td>
</tr>
<tr>
<td>Honesty Game (1)</td>
<td>Number of high dice rolls in honesty game</td>
</tr>
<tr>
<td>Management Experience (1)</td>
<td>Number of employees previously managed</td>
</tr>
<tr>
<td>Reading ability (1)</td>
<td>Score on reading test during survey</td>
</tr>
<tr>
<td>Writing ability (1)</td>
<td>Score on writing test during survey</td>
</tr>
<tr>
<td>Pro-Sociality/ Generosity (1)</td>
<td>Charitable donation in dictator game</td>
</tr>
<tr>
<td>Public Service Motivation (1)</td>
<td>Psychometric index measuring Public Service Motivation</td>
</tr>
<tr>
<td>Motivation (2)</td>
<td>Extrinsic/Intrinsic motivation scales</td>
</tr>
<tr>
<td>Big Five Personality Index (5)</td>
<td>Big Five Personality Index</td>
</tr>
<tr>
<td>Time Worked (1)</td>
<td>Hours worked per week</td>
</tr>
<tr>
<td>CHW Clients (1)</td>
<td>Total number of clients (as a CHW)</td>
</tr>
<tr>
<td>Peer Rating (3)</td>
<td>Rating of performance as a CHW, health knowledge, motivation</td>
</tr>
<tr>
<td>CHW Tasks (1)</td>
<td>Number of tasks carried out as a CHW</td>
</tr>
<tr>
<td>Medical Advice (1)</td>
<td>Frequency of giving advice on ailments (as a CHW)</td>
</tr>
<tr>
<td>DOTS provider (1)</td>
<td>Number of tuberculosis patients serving (as a CHW)</td>
</tr>
<tr>
<td>Institutional Deliveries (1)</td>
<td>Women brought to give birth at hospital (as a CHW)</td>
</tr>
<tr>
<td>CHW Performance (1)</td>
<td>Summary measure of performance (as a CHW)</td>
</tr>
<tr>
<td>Job Satisfaction as CHW (1)</td>
<td>5-point scale measuring their job satisfaction (as a CHW)</td>
</tr>
</tbody>
</table>

B.2 Validations of the Administrative Data on Health Services Delivery

One concern with the administrative data on CHW health services delivery is that the supervisors might find a way to manipulate it. I run three checks. First, I collected a second set of measures of supervisor performance from individuals overseeing the program but who were not part of hiring process. Supervisors could not manipulate this data since they did not know it was being collected. These individuals ranked the supervisors on 9 performance measures (e.g. overall performance, motivation) based on field visits and interactions with CHWs. They have no incentive to misreport
the performance of one supervisor relative to another; they knew this was only for research purposes and would have no impact on the supervisor. They also do not have access to supervisor-level administrative data on health services delivery, so their ratings are not mechanically related to the administrative data. Figure A.4 compares a PCA index of the 9 collected performance outcomes against the administrative data (average monthly change in functionality score for CHWs under the supervisor). The strong relationship indicate that the administrative data reflect ground reality rather than manipulation (column (1) of table A.6, \( p = 0.010 \)).

Second, I check how the administrative data compares to independent survey data. During the second survey round, CHWs were asked about how many deliveries they had assisted in each month for the last six months. Column (3) regresses the number of deliveries in the administrative data in a given month on the corresponding survey data. The recall periods differ slightly, so it is not surprising that the two do not align completely. However, the relationship is strong (\( p < 0.001 \)), arguing against major manipulation. Third, columns (2) and (4) shows that the relationship between the administrative and independent measures is not intermediated by the SPI of their supervisor. If manipulation explained my results, then we would expect the relationship between the administrative and independent data to be weaker for high SPI supervisors (negative interaction term).

Why would the supervisors not try to manipulate the data? There is little incentive for them to do so since this data does not affect their financial compensation. They are paid a fixed monthly amount, and it does not depend on any of these performance outcomes. The data would also be difficult to manipulate – for example, the data on deliveries assisted by the CHW is collected at the hospital level, where the supervisors have no jurisdiction.

### Table A.6: Validation of Administrative Data

<table>
<thead>
<tr>
<th></th>
<th>Functionality Score</th>
<th>Functionality Score</th>
<th>Deliveries (Admin)</th>
<th>Deliveries (Admin)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation Rating</td>
<td>0.117*** (0.0438)</td>
<td>0.108*** (0.0367)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation Rating X Supervisor SPI</td>
<td>-0.0357 (0.0280)</td>
<td></td>
<td>0.375*** (0.0259)</td>
<td>0.374*** (0.0260)</td>
</tr>
<tr>
<td>Deliveries (Survey)</td>
<td></td>
<td></td>
<td>0.0182</td>
<td></td>
</tr>
<tr>
<td>Deliveries (Survey) X Supervisor SPI</td>
<td></td>
<td></td>
<td></td>
<td>(0.0242)</td>
</tr>
</tbody>
</table>

Dependent mean .059 .059 2.1 2.1
Observations 917 917 3,005 3,005

*Notes: This table cross-checks the administrative data. Column (1) regresses administrative data for the CHWs supervised by a given supervisor against independent performance evaluations of those supervisors to check for manipulation. Column (3) reports estimates of the relationship between data on deliveries in a given month in administrative and survey data. Columns (1) and (2) tests whether supervisor SPI intermediates either relationship, which could be consistent with manipulation. * \( p < 0.10 \), ** \( p < 0.05 \), *** \( p < 0.01 \).
C Supplemental Analysis

C.1 Characterizing Hiring Decisions

At a high level, corrupt hiring systems may fall into three broad categories: (1) fully meritocratic, in which hiring decisions are made on the basis of merit, but the hiring committee demands a bribe from their preferred applicant; (2) fully non-meritocratic, in which hiring decisions are solely based on bribes and political connections; and (3) partially meritocratic, in which hiring decisions are made based on a combination of meritocratic and non-meritocratic elements. To categorize this hiring process, I use data on bribe offers, connections, and other characteristics of applicants. The hiring agent can be thought of as solving a discrete choice problem for each of the supervisor positions in which they select the applicant who maximizes their utility based on bribe offer, connections, and other characteristics. Under a fully meritocratic system, they would only value meritocratic characteristics, while under a fully non-meritocratic system, only bribes and political connections matter. More formally, the decision of a hiring agent to select a particular applicant \( i \) for a job \( j \) may be based on bribe offer made by the candidate \( b_{i,j} \), political connections of the candidate \( c_{i,j} \), a vector of candidate characteristics \( \alpha_{i,j} \), or an idiosyncratic preference for the applicant \( \epsilon_{i,j} \). The hiring agent can be modeled as having an objective function that determines their utility from hiring a particular applicant, as in equation (3).

\[
u_{i,j} = \pi b_{i,j} + \phi c_{i,j} + \theta \alpha_{i,j} + \epsilon_{i,j}\]

(3)

In a fully meritocratic system, \( \pi = \phi = 0 \). In a fully non-meritocratic system, only bribes and potentially political connections matter, and so \( \theta = 0 \).

The hiring agent will select the applicant who maximizes their utility, i.e. if \( w_j \) is the applicant hired in pool \( j \), then for all other applicants \( i' \) from pool \( j \), the utility from hiring \( w_j \) is greater than \( i' \), or \( \pi b_{w_j,j} + \phi c_{w_j,j} + \theta \alpha_{w_j,j} + \epsilon_{w_j,j} > \pi b_{i',j} + \phi c_{i',j} + \theta \alpha_{i',j} + \epsilon_{i',j} \). While many mechanisms might be used to elicit bribe offers (e.g. sealed bid or ascending price auctions), the exact system does not matter for this paper. For those making the hiring decisions, the bribe that matters for each applicant is the highest offer that she has made; in a sealed bid auction, this is submitted secretly, whereas in an ascending price auction, this is their last offer.

I use data on bribes, connections, and applicant characteristics in combination with the hiring agent’s revealed preference to estimate the weights that the hiring agent places on these factors. Assuming that \( \epsilon_{i,j} \) has a type-1 Gumbel distribution, I estimate \( \hat{\pi}, \hat{\phi} \) and \( \hat{\theta} \) to maximize the probability that the agent has the highest utility from selecting the set of observed hires, i.e. to maximize the likelihood expression:\(^2\)

\[
\prod_j \frac{e^{\pi b_{w_j,j} + \phi c_{w_j,j} + \theta \alpha_{w_j,j}}}{\sum_{i' \in I_j} e^{\pi b_{i',j} + \phi c_{i',j} + \theta \alpha_{i',j}}}\]

(4)

The first column of table A.7 estimates the likelihood expression with education as a relevant

\(^2\)This is not a logit regression of whether an applicant was hired on their characteristics and bribe offer. Instead, each likelihood term pertains to a position \( j \), and its value is the probability that under parameters \( \pi \) and \( \theta \), the applicant \( w_j \) who was hired for position \( j \) would be selected over each of the other applicants \( i' \) for that position.
characteristic of the applicant. The coefficients imply that use of a political connection increases the hiring agent’s utility over hiring that applicant by the same amount as offering an additional 4.68 salary-months of bribe, and that the hiring agent values each additional year of education at 1.64 salary-months of bribe. I reject the hypotheses that $\pi, \phi, \text{ or } \theta$ are equal to zero, and conclude that hiring decisions were partially meritocratic.

I generate two goodness of fit measures to evaluate the relative importance of these factors. The key input into these measures is $\hat{u}_{i,j} = \hat{\pi}b_{i,j}\hat{\phi}c_{i,j} + \hat{\theta}a_{i,j}$, the estimated utility of the hiring agent from hiring candidate $i$ given the estimated coefficients. The first goodness of fit measure is the percent of non-hires for whom $\hat{u}$ is lower than that of the winner of their competition, i.e. $\hat{u}_{w,j} \gt \hat{u}_i$ (pairwise % correct); if there were no idiosyncratic element $\epsilon_{i,j}$ and the true parameters were known, then hires would always have higher values of $u$ than non-hires. The second is the percent of hires who are correctly predicted, i.e. the predicted utility from selecting the hire is higher than selecting any of their competitors (winner % correct). Bribes, connections, and education correctly predicts the revealed preferred applicant in 88% of comparisons and predicts the winner in 86% of contests (column 1). The other columns test the predictive power of each variable individually. Bribe offers appear to be the main driver of the selection decision, given that they alone correctly predict 82% of the pairwise comparisons and 76% of the winners.

The two main take-aways are: (1) hiring was partially meritocratic; and (2) non-meritocratic factors explain most of the hiring decisions. Point estimates should be viewed as suggestive, as unobserved applicant characteristics may also be relevant. However, such unobserved characteristics would not change the main take-aways. Even if education were correlated with an unobserved quality such as public spiritedness on which hiring agents actually make their decisions, that would still be meritocratic. And even if bribe amounts are partially correlated with unobserved quality measures that were selected upon, the correlation would have to be implausibly high for fully meritocratic selection.

I carry out a number of robustness checks. First, in some clusters, the winning bribe offer was based on reports from other CHWs rather than the hire. Column (4) reruns the estimation with only clusters where the hire directly reported her bribe. The coefficients on bribes and political connections are statistically significant and similar in size, although the coefficient on education is no longer statistically significant, likely due to the reduced sample. Column (5) only uses clusters where the bribe of the hire was estimated based on informants. The consistency of these results with the main specification suggests that the other CHWs’ reports were accurate.

Second, some unsuccessful applicants refused to answer questions about bribes or claimed not to have offered a bribe. If these individuals did not offer bribes or would not have been competitive

---

43 To help identify the set of characteristics that might be relevant to the hiring decision, I surveyed non-applicants about reasons why they did not apply: the characteristics that they perceive themselves as lacking should be the set of characteristics that the hiring agent selects upon. I focus on the most frequently cited reasons: lacking financial resources to pay a bribe (28.1%) and not having enough education (48.4%).

44 I have also tested other plausible predictors of hiring, including age, past management experience, wealth, past performance as a health worker, health knowledge, problem-solving abilities and psychometric measures, but none are statistically significant predictors of hiring.

45 For example, wealth is the best predictor of bribe size, but still only explains around 15% of the variation in bribes; selection on unobservable factors would have to be implausibly stronger than on wealth to explain the selection on bribes.
applicants, then their omission will not bias estimates.\footnote{Parameter estimates depend mostly on individuals on the threshold of selection, since marginal changes in parameter values affect the probability that those individuals are selected. For those who are far from the threshold of being selected (either because they did not offer bribes or their offers were noncompetitive), changes in the parameter values have little effect on the overall likelihood expression, and so their inclusion makes little difference.} However, it could be a problem if some unsuccessful applicants offered larger bribes than the selected applicants, but did not report them. This is unlikely for three reasons. First, unsuccessful applicants who did not report offering a bribe are poorer, at the 51.1\textsuperscript{st} percentile of wealth, than those who said they offered bribes (57.5\textsuperscript{th} percentile). Wealth is the most important determinant of bribe offers, so this is fairly indicative. Second, appendix table A.8 re-runs the estimation with imputed bribes based on observable characteristics for applicants with no reported bribe. This has no effect on the estimates (column 1), and findings are robust to inflating the imputed bribe values considerably (column 2). Column (3) imputes bribes for the supervisors who did not report a bribe (but not for unsuccessful applicants), and column (4) imputes bribes for anyone who did not report bribes; results are again similar. Third, I examine contests in which only a small number of candidates do not report bribe offers, i.e. the set of bribe offers is likely more complete. Column 5 includes competitions in which two or fewer applicants do not report bribe offers, while column 6 includes only competitions in which one or fewer applicants do not report bribe offers. While the coefficients on connections and education shift, potentially due to the smaller samples, the steadiness of the coefficient on bribe demonstrates that bribe size is important in hiring decisions.

Table A.7: Hiring decisions

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>All</th>
<th>All</th>
<th>All</th>
<th>Primary Reports Only</th>
<th>Secondary Reports Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bribe Amount</td>
<td>1.00***</td>
<td>0.89***</td>
<td></td>
<td></td>
<td>0.95***</td>
<td>1.62***</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.10)</td>
<td></td>
<td></td>
<td>(0.19)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>Political Connections</td>
<td>4.68***</td>
<td>5.22***</td>
<td></td>
<td></td>
<td>3.71*</td>
<td>9.64*</td>
</tr>
<tr>
<td></td>
<td>(1.20)</td>
<td>(1.13)</td>
<td></td>
<td></td>
<td>(1.89)</td>
<td>(1.96)</td>
</tr>
<tr>
<td>Education</td>
<td>1.64***</td>
<td>1.32***</td>
<td></td>
<td></td>
<td>0.56</td>
<td>5.29**</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.21)</td>
<td></td>
<td></td>
<td>(0.46)</td>
<td>(0.38)</td>
</tr>
<tr>
<td>Pairwise % Correct</td>
<td>0.88</td>
<td>0.82</td>
<td>0.56</td>
<td>0.43</td>
<td>0.81</td>
<td>0.96</td>
</tr>
<tr>
<td>Winner % Correct</td>
<td>0.86</td>
<td>0.76</td>
<td>0.30</td>
<td>0.30</td>
<td>0.81</td>
<td>0.89</td>
</tr>
<tr>
<td>Observations</td>
<td>191</td>
<td>191</td>
<td>191</td>
<td>191</td>
<td>89</td>
<td>102</td>
</tr>
</tbody>
</table>

Notes: This table evaluates the factors that enter into hiring decisions. The coefficient values are equal to the increase in utility of the hiring agent (measured in salary months of bribe) from hiring a candidate with one unit higher value of the relevant independent variable. Column (1) includes all three predictors in the estimation, while columns (2)-(4) examine each element separately. Column (5) restricts to clusters in which the supervisor told surveyors their bribe payment, while column (6) is for clusters where they did not. Coefficients are estimated via maximum likelihood and standard errors are based on the likelihood function. * \( p < 0.10 \), ** \( p < 0.05 \), *** \( p < 0.01 \)
Table A.8: Hiring decisions (robustness checks)

<table>
<thead>
<tr>
<th></th>
<th>Non-Hire</th>
<th>Non-Hire</th>
<th>Supervisor</th>
<th>All</th>
<th>Missing</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Imputed</td>
<td>Imputed</td>
<td>Imputed x1.33</td>
<td>Imputed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bribe Amount</td>
<td>0.85***</td>
<td>0.26**</td>
<td>0.64***</td>
<td>0.56***</td>
<td>1.41***</td>
<td>2.02**</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.09)</td>
<td>(0.12)</td>
<td>(0.11)</td>
<td>(0.13)</td>
<td>(0.23)</td>
</tr>
<tr>
<td>Political Connections</td>
<td>6.45***</td>
<td>5.59***</td>
<td>4.69***</td>
<td>6.14***</td>
<td>2.12</td>
<td>−1.27</td>
</tr>
<tr>
<td></td>
<td>(1.08)</td>
<td>(1.08)</td>
<td>(1.13)</td>
<td>(1.27)</td>
<td>(1.35)</td>
<td>(2.13)</td>
</tr>
<tr>
<td>Education</td>
<td>1.14***</td>
<td>1.14***</td>
<td>1.46***</td>
<td>1.13***</td>
<td>2.63**</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.24)</td>
<td>(0.26)</td>
<td>(0.26)</td>
<td>(0.35)</td>
<td>(0.56)</td>
</tr>
<tr>
<td>Pairwise % Correct</td>
<td>0.86</td>
<td>0.77</td>
<td>0.80</td>
<td>0.82</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td>Winner % Correct</td>
<td>0.70</td>
<td>0.52</td>
<td>0.73</td>
<td>0.57</td>
<td>0.90</td>
<td>0.93</td>
</tr>
<tr>
<td>Observations</td>
<td>322</td>
<td>322</td>
<td>192</td>
<td>322</td>
<td>109</td>
<td>65</td>
</tr>
</tbody>
</table>

Notes: This table contains robustness checks for table A.7. Columns (1)-(2) impute bribe offers based on observable characteristics for non-hires who say that they did not offer a bribe, with column (2) inflating that imputed value by 33%. Column (3) imputes bribes for supervisors who did not say their bribe amount, while column (4) imputes bribes for anyone who did not state a bribe amount. Columns (5) and (6) includes only competitions where there is less missing data (with two and one missing bribe offer respectively). Coefficients are estimated via maximum likelihood and standard errors are based on the likelihood function. ∗ p < 0.10, ∗∗ p < 0.05, ∗∗∗ p < 0.01

C.2 Auxiliary Tests on the Quality and Effect of Supervisors

C.2.1 Differences-in-Differences Estimates

In the main text, the paper uses differences regressions with administrative data from after the supervisor is hired. One concern is that the differences specification misses the immediate effect of the introduction of supervisors on CHW performance since it only looks in changes in the post-hiring period. I would ideally run a differences-in-differences specification with data from before and after the supervisor start date, but was not able to access administrative data from prior to the start of the supervisors. However, I did collect data on some indicators of CHW performance during both rounds of surveys. One survey occurred prior to the supervisor training, and the other was conducted after supervisors had worked for six months, so I can run a differences-in-differences regression to test for immediate performance improvements for CHWs under higher SPI supervisors. The outcomes measured in both surveys were: 1) number of pregnant women whom the CHW currently assists; 2) hours worked per week; 3) households mobilized for immunization in the previous month; and 4) institutional deliveries assisted.47

Table A.9 finds that the short-run impact of the supervisors is limited. While higher SPI supervisors do increase the hours worked by their CHWs (p = 0.018), this does not manifest in service delivery gains within six months. Instead, it appears that it takes some time for supervisors to generate real change in their workforce, meaning that the use of a differences specification should be appropriate in determining the effect of supervisors. If anything, the differences regression would underestimate the effect of supervisors given that supervisors do immediately increase the hours that their subordinates work.

47For the first three outcomes, I run a simple differences-in-differences specification of $y_{ict} = \alpha + \beta_1 post_t + \beta_2 SPI_{supervisor_{ic}} + \beta_3 post*SPI_{supervisor_{ict}} + \epsilon_{ict}$. The number of deliveries is observed monthly, so I run the regression $y_{ict} = \alpha + \beta post*SPI_{supervisor_{ict}} + \gamma_c + \phi_t + \epsilon_{ict}$ where $\gamma_c$ and $\phi_t$ are CHW and month fixed effects.
Table A.9: Differences-in-differences regression for CHW outcomes

<table>
<thead>
<tr>
<th></th>
<th>Pregnant Clients</th>
<th>Hours Worked</th>
<th>Immunization Mobilized</th>
<th>Deliveries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post X Supervisor SPI</td>
<td>-0.017</td>
<td>0.87**</td>
<td>0.36</td>
<td>-0.056</td>
</tr>
<tr>
<td></td>
<td>(0.274)</td>
<td>(0.357)</td>
<td>(0.873)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>Observations</td>
<td>1276</td>
<td>1277</td>
<td>1224</td>
<td>6400</td>
</tr>
</tbody>
</table>

Notes: This table reports results from a differences in differences regression of service delivery outcomes on a post period dummy, supervisor SPI and the interaction of the post period dummy and supervisor SPI. Column (4) includes month and CHW fixed effects. Standard errors are clustered at supervisor level.

C.2.2 Additional Quality Measures

One concern with the data used in the paper is that it was collected after hiring decisions had been announced. It is possible that supervisors may have gained skills after being hired and prior to data collection that complicate the comparisons between them and non-hires. This will not be an issue for characteristics that do not vary over this time frame (e.g. education, which all have completed, or cognitive ability). However, reading/writing ability, health knowledge, and performance as a CHW (in the period between their hiring and starting their work as supervisors) may have responded to their hiring. I take two approaches to address this concern using data sources from before the hiring from two of the eight study regions. The first to use the data to test whether there is indeed a differential change for supervisors relative to non-supervisors over the period between the supervisors being hired and my main data collection. The second is to test for relative quality of supervisors prior to hiring.

The first data set consists of tests of reading ability and health knowledge that were administered to CHWs approximately a year before the hiring of supervisors was announced, while the second is monthly-level data on CHW salaries beginning approximately a year and a half prior to the announcement of hiring of supervisors (see figure A.1 for a visual representation of the timing). Table A.10 uses the first data set to test for differential changes in reading ability and health knowledge for supervisors relative to the rest of the population. I use the two differences-in-differences specifications in equations 5 and 6, with data from the pre-hiring period \((t = 0)\) and post-hiring data collection \((t = 1)\). \(y_{ct}\) is the outcome for CHW \(c\) in period \(t\), \(post_{ct}\) is equal to a dummy for if the data comes from the post-hiring period, \(supervisor_{ct}\) is a dummy for whether the CHW is hired as a supervisor, and \(\phi_c\) is a CHW fixed effect. \(\beta_4\) tests for differential changes over time for supervisors, and \(\beta_3\) tests for differences between supervisors and non-hires in the pre-hiring period. For both health knowledge and reading test scores, I find that the supervisors had much better scores in the pre-period and no evidence of a differential improvement in the post period. This supports the use of post-hiring data for reading/writing ability and health knowledge.

\[
y_{ct} = \beta_1 + \beta_2 post_{ct} + \beta_3 supervisor_{ct} + \beta_4 supervisor_{ct} post_{ct} + \epsilon_{ct} \tag{5}
\]

\[
y_{ct} = \beta_1 + \beta_2 post_{ct} + \beta_4 supervisor_{ct} post_{ct} + \phi_c + \epsilon_{ct} \tag{6}
\]
The second approach examines the quality of supervisors relative to non-applicants and unsuccessful applicants in the pre-hiring period. I first look at salaries in the two years prior to hiring: CHWs are paid based on their delivery of specific health services in their village, and so the salary data reflects their performance in a given month, particularly in bringing women to give birth at a facility rather than at home, serving as a tuberculosis treatment provider, immunizing children, and promoting adoption of long-acting contraception. Column (1) of Table A.11 finds that the average earnings of CHWs selected as supervisors were 28% higher than those of unsuccessful applicants ($p = 0.029$), and 45.9% higher than those did not apply ($p = 0.007$) prior to the announcement of hiring for the supervisor position. This can be also seen in the top panel of figure A.5, which graphs a moving average of the salaries of non-applicants, applicants, and selected supervisors over time. Consistent with the data in the paper (table I), past performance of those selected as supervisors was better than that of those who were not selected.\footnote{The bottom panel of figure A.5 plots the salaries of actual hires against counterfactual hires. The sample sizes are small, but the counterfactual and actual hires again are comparable.}

Second, the earlier referenced health knowledge test was split into three types of questions – 12 questions on danger signs for mothers during pregnancy, 12 questions on danger signs for infants/mothers after delivery, and 18 questions on general health knowledge. The below table finds the same pattern as in table I: those who applied for the job as a supervisor had consistently better health knowledge (note that supervisors are among the set who applied for the job). In one of three categories, the difference between supervisors and unsuccessful applicants is statistically significant, so as in table I, it appears that supervisor’s health knowledge is weakly better than unsuccessful applicants in the pre-hiring period.

\begin{table}[h]
\centering
\begin{tabular}{cccc}
\hline
 & (1) & (2) & (3) & (4) \\
 & Health Knowledge & Health Knowledge & Reading Ability & Reading Ability \\
Supervisor & 0.243*** & 0.784** & & \\
 & (0.0772) & (0.387) & & \\
Supervisor X Post & 0.0467 & 0.0467 & 0.278 & 0.278 \\
 & (0.109) & (0.0849) & (0.547) & (0.206) \\
Observations & 464 & 464 & 464 & 464 \\
Dependent Mean & 0.488 & 0.488 & 4.026 & 4.026 \\
CHW FE\$s & No & Yes & No & Yes \\
\hline
\end{tabular}
\caption{Differential Changes in Reading and Health Knowledge After Hiring}
\textit{Notes:} All specifications also include a dummy variable for whether the data come from the post-hiring period. Columns (1) and (2) test for differential changes in health knowledge for supervisors and non-supervisors, while columns (3) and (4) test for such changes in reading ability. Columns (2) and (4) include CHW fixed effects to improve the precision of the estimates. Standard errors are clustered at the CHW level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ 
\end{table}
Table A.11: Comparison Across Hires, Unsuccessful Applicants, and Non-Applicants

<table>
<thead>
<tr>
<th>Health Knowledge Test</th>
<th>Monthly Earnings</th>
<th>Pregnancy Health</th>
<th>Newborn Health</th>
<th>General Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hired (=1)</td>
<td>0.285**</td>
<td>1.171***</td>
<td>-0.214</td>
<td>-0.0571</td>
</tr>
<tr>
<td></td>
<td>(0.130)</td>
<td>(0.404)</td>
<td>(0.212)</td>
<td>(0.429)</td>
</tr>
<tr>
<td>Applied (=1)</td>
<td>0.174***</td>
<td>0.557***</td>
<td>0.104</td>
<td>0.895***</td>
</tr>
<tr>
<td></td>
<td>(0.0639)</td>
<td>(0.200)</td>
<td>(0.105)</td>
<td>(0.211)</td>
</tr>
<tr>
<td>Dependent mean</td>
<td>1</td>
<td>8.7</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>Observations</td>
<td>236</td>
<td>236</td>
<td>236</td>
<td>236</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates of the relationship between various measures of quality and whether a CHW was selected as a supervisor, using measures of quality collected prior to the hiring of supervisors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 
Figure A.5: Salary Payments Prior to Hiring of Supervisors

(a) Salary of Hires, Unsuccessful Applicants, and Non-Applicants

(b) Salary of Actual and Counterfactual Hires

Notes: This figure plots a three month moving average of the salaries of groups of CHWs before and after the hiring of supervisors, but before the CHWs hired as supervisors began working in their new role. CHWs are paid on a piece rate, and so this reflects performance as a health worker. Sub-figure A.5a divides the community health workers into those who did not apply for the job as supervisor, those who applied and were not selected, and those who were hired. Sub-figure A.5b divides the community health workers into those who were hired as supervisors and those who would have been hired under three counterfactual hiring systems.