Service Delivery, Corruption, and Information Flows in Bureaucracies: Evidence from the Bangladesh Civil Service

Martin Mattsson*

Job Market Paper

17th November 2020

Click here for latest version

Abstract

Government bureaucracies in low- and middle-income countries often suffer from corruption and slow public service delivery. Can an information system – providing information about delays to the responsible bureaucrats and their supervisors – reduce delays? Paying bribes for faster service delivery is a common form of corruption, but does improving average processing times reduce bribes? To answer these questions, I conduct a large-scale field experiment over 16 months with the Bangladesh Civil Service. I send monthly scorecards measuring delays in service delivery to government officials and their supervisors. The scorecards increase services delivered on time by 11% but do not reduce bribes. Instead, the scorecards *increase* bribes for high-performing bureaucrats. These results are inconsistent with existing theories suggesting that speeding up service delivery reduces bribes. I propose a model where bureaucrats' shame or reputational concerns constrain corruption. When bureaucrats' reputation improves through positive performance feedback, this constraint is relaxed, and bribes increase. Overall, my study shows that improving information within bureaucracies can change bureaucrats' behavior, even without explicit incentives. However, positive performance feedback can have negative spillovers on bureaucrats' performance across different behaviors.

*Yale University, martin.mattsson@yale.edu. I would like to thank Gaurav Chiplunkar, Anir Chowdhury, Andrew Foster, Eduardo Fraga, Sahana Ghosh, Marina Halac, Ashraful Haque, Enamul Haque, Daniel Keniston, Ro'ee Levy, Imran Matin, Mushfiq Mobarak, Farria Naeem, Rohini Pande, Nick Ryan, Mark Rosenzweig, Jeff Weaver, Jaya Wen, Fabrizio Zilibotti, and numerous seminar participants for helpful comments and suggestions. I also thank Mahzabin Khan and Ashraf Mian for excellent research assistance, as well as IPA Bangladesh for outstanding research support. A randomized controlled trial registry and pre-analysis plan are available at: www.socialscienceregistry.org/trials/3232. This work was supported by the JPAL Governance Initiative (GR-0861), the International Growth Centre (31422), the Yale Economic Growth Center, the MacMillan Center for International and Area Studies, the Weiss Family Fund, and the Sylff fellowship.

1 Introduction

The state's capacity to implement its policies, secure property rights, and provide basic public services is paramount for economic development. To have this capacity, the state needs a functioning bureaucracy of government officials motivated to carry out their tasks. For career civil servants, compressed wage structures, secure employment, and opportunities for rent extraction through corruption often lead to weak or counterproductive incentives, especially in low- and middle-income countries. While explicit incentive structures, such as pay-for-performance contracts, can change the behavior of government officials, they are often hard to implement without unintended consequences (Finan, Olken, and Pande, 2017). Furthermore, political constraints often prevent the introduction of explicit incentive structures altogether. However, the lack of explicit incentives does not mean that civil servants have no incentives. Supervisors in government bureaucracies often influence future postings and career paths of lower-level bureaucrats, which can be a strong motivating factor for civil servants (Khan, Khwaja, and Olken, 2019). Furthermore, bureaucrats may have strong intrinsic motivations to perform their jobs well (Banuri and Keefer, 2013; Cowley and Smith, 2014).

Providing better information flows within bureaucracies about individual officials' performance may improve existing incentives by allowing supervisors to align postings and promotions more closely with job performance. Regular feedback may also increase officials' intrinsic motivation by making their own performance more salient to themselves. Furthermore, the flexible interpretation of information that is not directly tied to explicit incentives may avoid some of the common pitfalls of explicit incentives structures such as the neglect of tasks not measured by the performance indicators and opposition from individuals within the organization leading to poor implementation (Banerjee, Chattopadhyay, Duflo, Keniston, and Singh, 2020). Historically, high-frequency information on bureaucrat performance has often been expensive to collect, but e-governance systems can substantially reduce this cost and increase the data quality (Singh, 2020). As low- and middle-income countries have expanded their digital capabilities, this has created new opportunities for improved information systems in the management of government officials.

This paper focuses on the processing time of applications for changes to government land records in Bangladesh. An update to the government records has to be made every time a parcel of land changes owners and is necessary for the issuance of a land title to the new owner. Updated land records and land titles are essential for individuals to have secure property rights over land. Land disputes are one of the most severe legal problem in Bangladesh, with 29% of adults having faced a land dispute in the past four years (Hague Institute for Innovation of Law, 2018). Slow public service delivery is also a significant problem in Bangladesh. For example, only 56% of land record change applications in my control group are processed within a 45 working day time limit mandated by the government. Furthermore, faster service provision is a commonly stated reason for bribe payments, suggesting that slow service delivery on average may cause corruption as some firms and citizens pay bribes to avoid having to wait for their services.[1]

In an experiment with the Bangladesh Civil Service, I provide information regarding junior civil servants' performance using monthly scorecards sent to the civil servants themselves and their supervisors. The scorecards are designed to reduce delays in the processing of applications for land record changes and are based on data from an e-governance system. There are two performance indicators shown on the scorecards: the number of applications processed within the official time limit of 45 working days and the number of applications pending beyond that limit. The scorecards also show the bureaucrats' relative performance on these indicators, compared to all other bureaucrats in the experiment. The intervention is randomized at the level of the land office, and there is only one civil servant per office. The experiment was carried out at a large scale and involve 311 land offices (59% of all land offices in Bangladesh), which serve a population of approximately 95 million people.

The scorecards had a meaningful effect on bureaucrats' behavior. Using administrative data on more than a million applications, I estimate that the scorecards increase the share of applications processed within the time limit by 6 percentage points or 11%. The effect starts almost immediately after the scorecards are first sent out and is present for the 16 month period of the experiment. The scorecards also decreased the average processing time of applications by 13% and the applicants' visits to government offices by 12%. The effects are almost entirely driven by bureaucrats in offices with a below-median performance at baseline, improving their performance. This result shows that improving the information flows within a bureaucracy can change bureaucrats' behavior, even without explicit incentive structures.

Since the scorecards were sent to both the bureaucrats and their supervisors, there might be two different mechanisms for the effect on behavior. First, the bureaucrats may care about their reputation among their supervisors, potentially because of the influence the supervisors have over their careers. Second, the scorecards may also change bureaucrats' behavior by causing a sense of shame or pride through making their absolute and relative performance more salient to the bureaucrats themselves. For

[1] Among households in Bangladesh reporting having paid a bribe for a public service, 23% stated that "timely service" was one of the reasons for paying the bribe (Transparency International Bangladesh, 2018).

ease of exposition, I will refer to these two concerns as *reputational concerns*. While I cannot distinguish between these two mechanisms, I use a variation of the scorecard to test if peer effects from having the performance information shared among bureaucrats at the same level within the bureaucracy can motivate bureaucrats further than having the information shared only with the supervisors. I find no evidence of meaningful peer effects beyond the effect of the standard scorecards.

Some existing theories of corruption suggest that the average speed of public service delivery is causally and negatively related to bribes since, when average processing times are long, some applicants pay to avoid having to wait (Leff, 1964; Rose-Ackerman, 1978; Kaufmann and Wei, 1999). I conduct a survey among applicants and use the experimental variation in processing times to test theories of how they are related to corruption. Overall, the scorecards did not decrease bribe payments. The point estimate of the effect is an increase of BDT 1,046 (~USD 12) or 17%, and the lower bound of the 95% confidence interval is a decrease of 3%.[2] The increase comes from a positive effect on the bribe amounts reported (intensive margin) with no effect on the fraction of applicants reporting bribes (extensive margin). Using an experimental information intervention among surveyed applicants, I rule out that the lack of a decrease in bribes is due to the information about the improved processing times not yet having disseminated among applicants.

The positive effect of the scorecards on bribe payments is concentrated among the offices that were over-performing at baseline, i.e., the offices for which the scorecards have no effect on processing times. In the under-performing offices, where scorecards improve processing times, they do not affect bribes. This is inconsistent with a causal relationship between average processing times and bribes since processing times can improve without bribes changing, and bribes can increase without processing times changing.

I propose a model in which bureaucrats trade-off reputational concerns, bribe money, and the utility cost of effort. The bureaucrats' reputation is determined by their visible job performance along two dimensions, delays and bribe extraction, which are only imperfectly observable to supervisors. The scorecards increase the visibility of delays and thereby make them more important for reputation. For under-performing bureaucrats, this also means that the scorecards decrease their reputation. Therefore, the model predicts that under-performing bureaucrats reduce delays by providing more effort. The model also predicts that when the scorecards highlight the already good performance of over-performing bureaucrats, this relaxes their reputation constraint, allowing them to increase bribes. Furthermore, the

[2] Throughout the paper, I use a USD/BDT exchange rate of 84.3, the average exchange rate during the experiment.

model is consistent with the result that the scorecards do not affect delays for over-performing bureaucrats. For them, more visible delays increase incentives to avoid delays (substitution effect), but increased reputation has made the marginal importance of reputation smaller (income effect), so the overall effect is ambiguous.

This paper contributes to four strands of literature. First, it contributes to the literature on how incentives shape bureaucratic performance. There is an extensive literature on both monetary and non-monetary explicit incentives (e.g., Ashraf, Bandiera, and Jack, 2014; Khan, Khwaja, and Olken, 2016; Khan et al., 2019). This paper contributes to the growing literature on the effects of information flows within government bureaucracies (Dodge, Neggers, Pande, and Moore, 2018; Muralidharan, Niehaus, Sukhtankar, and Weaver, 2020; Dal Bó, Finan, Li, and Schechter, 2019; Callen, Gulzar, Hasanain, Khan, and Rezaee, 2020; Banerjee et al., 2020). In particular, I show that information flows about individual civil servants' performance can improve public service delivery even without explicit incentives and that this effect is persistent over time. This could be due to long-term career concerns of bureaucrats (Niehaus and Sukhtankar, 2013a; Bertrand, Burgess, Chawla, and Xu, 2020) or a sense of shame or pride internal to the bureaucrats themselves (Allcott, 2011; Dustan, Maldonado, and Hernandez-Agramonte, 2018). However, I find no evidence that reputational concerns *among* bureaucrats at the same level in the organizational hierarchy are a substantial motivating factor (Mas and Moretti, 2009; Cornelissen, Dustmann, and Schönberg, 2017). My model suggests that reputational concerns provide incentives for performance and limit the amounts of bribes collected by bureaucrats. However, the model also shows how improving the relative reputation of individual bureaucrats can lead them to perform worse and be more corrupt.

Second, the paper provides empirical evidence on the connection between corruption and the speed of public service delivery, or more generally, red tape. Slow service delivery is positively associated with corruption (Kaufmann and Wei, 1999; Freund, Hallward-Driemeier, and Rijkers, 2016), and applicants may have to pay bribes to increase processing speed for services (Bertrand, Djankov, Hanna, and Mullainathan, 2007). In the mainly theoretical literature on why the speed of service delivery and corruption are associated, different models lead to drastically different policy conclusions. One view is that corruption allows firms and individuals to circumvent excessively onerous bureaucratic hurdles (Leff, 1964; Huntington, 1968). In this view, rooting out corruption would decrease the speed of service delivery and increase inefficiencies of excessive bureaucratic control. An opposing view is that corruption is the driver of red tape and delays in public services, as making the de-jure regulation more onerous allows government officials to extract more bribes (Myrdal, 1968; Rose-Ackerman, 1978; Kaufmann and Wei, 1999).[3] According to this view, we could improve service delivery by eliminating corruption. According to both views, we could reduce corruption by providing services with fewer delays to everyone. I contribute to the literature by showing that, in this context, increasing the average speed of service delivery does not decrease bribe payments and that there is no evidence of a causal relationship between the average speed of service delivery and bribes.

Third, the paper contributes to the literature on the determinants of bribe amounts. In some settings, bribe payers' outside option and ability to pay constrain bribe amounts (Svensson, 2003; Bai, Jayachandran, Malesky, and Olken, 2019), potentially leaving little room for applicant complaints or monitoring to reduce corruption (Niehaus and Sukhtankar, 2013b). In other settings, monitoring has been effective in reducing corruption (Reinikka and Svensson, 2005; Olken, 2007). I show that, in this context, individual bureaucrats *can* increase bribes and that bribes do not just reflect the difference between the official fee and the applicants' willingness or ability to pay for the service. Instead, my model highlights how bureaucrats' reputational concerns constrain bribes, explaining why bribes are substantially below applicants' willingness to pay for the service.

Finally, the paper is related to the literature on the effects of e-governance in settings with low government capacity. In some cases, e-governance systems have improved government efficiency and reduced corruption (Banerjee, Duflo, Imbert, Mathew, and Pande, 2020; Lewis-Faupel, Neggers, Olken, and Pande, 2016). While in others, they have not had substantial benefits and wasted scarce government resources (World Development Report, 2016). This paper does not evaluate an e-governance system as a whole. Instead, it provides evidence on the untapped potential in the data that e-governance systems generate for the management of government officials.

The rest of the paper is organized as follows. Section 2 describes the context, experimental interventions, and data. Section 3 describes the empirical strategy used to analyze the experiment. Section 4 presents the estimated effects of the scorecards on processing times and bribes. Section 5 discusses how the results relates to existing theories of the relationship between the speed of service delivery and corruption. Section 6 proposes a model of bureaucratic behavior explaining the results. Section 7 concludes by discussing policy implications.

[3] In Banerjee (1997), both corruption and red tape emerge from the nature of public service provision in low-income countries due to a principal-agent problem between the government and its bureaucrats. The experimental results can neither reject nor confirm this model.

2 Context, Experimental Intervention, and Data

The context of this study is land record changes in Bangladesh, and specifically the time it takes to process applications for such changes. Maintaining an updated record of land ownership is crucial for secure property rights, and globally it is an example of a public service that is almost exclusively provided by the state. More generally, the timely provision of public services is an important aspect of government capacity. The speed of public service provision is a key determinant of a country's score in the World Bank's annual *Doing Business* report. Timely public service provision and policy implementation has been shown to be positively associated with poverty reduction (Djankov, Georgieva, and Ramalho, 2018), trade (Djankov, Freund, and Pham, 2010), entrepreneurship (Klapper, Laeven, and Rajan, 2006), and economic output (Nicoletti and Scarpetta, 2003; Djankov, McLiesh, and Ramalho, 2006). Several countries, such as India and Russia, have explicitly stated goals to reach a certain Doing Business ranking, showing the importance that governments in low- and middle-income countries place on increasing the speed of public service delivery.[4]

The scorecard intervention is made possible by a recently implemented e-governance system that bureaucrats in Bangladesh use to process applications for land record changes. Appendix Figure A1 shows how governments in low- and middle-income countries have expanded their digital capacity compared to high-income countries in four important areas of governance. The figure shows that that public services provided using e-governance systems are now commonplace, if not the norm, even outside high-income countries.

2.1 Land record changes in Bangladesh

When a parcel of land changes owners in Bangladesh, either through sale or inheritance, the official land record has to be changed and a new record of rights issued to the new owner. Land record changes (called "mutations" in Bangladesh) are conducted by civil servants holding the position of Assistant Commissioner Land (ACL), whom I am referring to as the *bureaucrats*. ACL is a junior position in the Bangladesh Administrative Service, the elite cadre of the Bangladesh Civil Service. Each ACL heads a sub-district (*Upazila*) land office. The ACL is directly supervised by an Upazila Nirbahi Officer (UNO), the most senior civil servant at the sub-district level. The UNO is then supervised by a Deputy Commis-

[4]India aims to be in the top 50 (https://www.livemint.com/Politics/D8U9SSxwJ741OH7CxYlZeO/India-unlikely-to-see-significant-rise-in-Doing-Business-ran.html) while Russia aims to be in the top 20 (https://russiabusinesstoday.com/economy/russia-advances-in-doing-business-ranking-but-fails-to-enter-top-20/).

sioner (DC), the most senior bureaucrat at the district level. The UNO has substantial power over the ACL's future career through an Annual Confidential Report regarding the performance of the ACL that the UNO submits to the Ministry of Public Administration. I am referring to the UNOs and DCs as the *supervisors*.

A bureaucrat typically holds the position of ACL for one to two years and when an ACL is transferred, it is often to a different position within the bureaucracy. For example, of the 615 ACLs I observe in my administrative data, only 10% held the position of ACL in more than one land office.

The de-jure process for making a land record change is visually represented in Appendix Figure A2. To make a land record change, the new owner must apply for such a change at the sub-district land office where the land is located. Hence, there is no competition between land offices for applicants. The application is then inspected by the office staff, who verify that the application has the required documents. The application is then sent to the local (*Union Parishad*) land office of the area where the land is located. The local land office is the lowest tier of land offices and is staffed by a Land Office Assistant who verifies the applicant's claim to the land by meeting with the applicant and visually inspecting the land. The Land Office Assistant then writes a recommendation on whether to accept or reject the application to the sub-district land office. The application is then verified against the existing government land record. Finally, a meeting is held between the ACL and the applicant where, the application is formally approved. The applicant then pays the official fee of BDT 1,150 (USD ~14) for the issuance of the new record of rights. When the applicant has paid the fee, the new record of rights is issued and given to the applicant. The sub-district land office also changes the official government land record to reflect the new ownership. The Government has mandated that land record changes should take no more than 45 working days, but in practice delays beyond this time limit are common. In my data, only 56% of applications in the control group were processed within the time limit and the average processing time among processed applications was 52 working days.

2.1.1 Bureaucrats' discretionary powers and corruption

In practice, it is common for applicants to also pay bribes beyond the official fee to get their application processed. Figure 1 shows that among the applicants in my survey, the average estimated bribe for a typical applicant was BTD 6,731 (~USD 80).[5] Appendix Figure A3 shows that when asked, the most common response to the question of why a bribe was paid is akin to "to get the work done" (39%), the

[5] Appendix Figure A8 shows that this estimate is similar to an estimate by Transparency International Bangladesh of the average bribe paid for a land record change.

second most common response is akin to "to avoid hassle" (39%), and the third most common is akin to "for faster processing" (10%). This highlights that the bureaucrat has decision making power over the application along two dimensions. First, they can decide whether to accept or reject the application. Second, they can take actions to speed up or slow down the application as well as create more or less hassle for the applicant. Figure 1 shows that the average stated valuation of getting the record of rights is BDT 1,594,664 (~USD 18,917), almost as high as the average estimated market value of the land itself. These valuations are more than two orders of magnitude larger than even the highest estimate of the average bribe payments.

On average, the applicants in my survey state that their willingness to pay for having their application processed within seven days (the shortest reasonable processing time) is BDT 2,207 (~USD 26). Since this number is substantially lower than the average bribe paid by those reporting a non-zero bribe and the average estimated typical bribe, it is clear that applicants are not just paying for faster processing. Most likely they are also paying for getting the approval.

2.2 E-governance system for land record changes

In February 2017, a new e-governance system for land record changes was introduced, with the goal of simplifying the process of land record changes for both the applicants and the civil servants processing the applications. The system was gradually implemented in sub-district and local land offices. As the e-governance system had recently been implemented at the time of the experiment, not all applications were processed using the e-governance system even in the sub-district offices where it had been installed. The main reason for this was that not all local land offices within the sub-district had had the e-governance system installed.[6]

The e-governance system generates administrative data on each application made in the system. Specifically, this administrative data can be used to assess the adherence to the rule that all applications should be processed within 45 working days. However, until the start of the experiment, this data was not presented in a format enabling evaluation of the degree of adherence to this rule or the performance of specific sub-district land offices or ACLs.

[6] Other reasons cited for using the paper-based system were problems with internet connectivity, new officials not yet trained in using the e-governance system, and temporary problems with the e-governance server.

2.3 Experimental intervention: Performance scorecards

Together with the Government of Bangladesh, I designed a monthly performance scorecard addressed to the ACL and sent to randomly selected sub-district land offices, as well as to the offices of the UNO and the DC, the ACL's two direct superiors. The scorecard is intended to decrease delays in application processing for land record changes. Appendix Figure A4 presents an example of a performance scorecard.

The scorecard evaluates the ACL's performance using two performance indicators. The first indicator is the number of applications disposed within 45 working days in the past month, where a higher number indicates a better performance. The second indicator is the number of applications pending beyond 45 working days at the end of the month, where a lower number indicates a better performance. The scorecard shows both these numbers as well as the average numbers for all sub-district land offices in the experiment. The scorecard also provides the office's percentile ranking for each indicator, with a short sentence reflecting the performance. Finally, to make the score easily understandable and more salient, a thumbs-up symbol is put next to percentile rankings between the 60th and the 100th percentile, while a thumbs-down symbol is put next to percentile rankings from the 0th percentile to the 40th percentile. Two versions of the scorecard, one in English and one in Bengali, were sent out in the first two weeks of each month with information based on the previous calendar month's e-governance data. Offices in the treatment group were not informed that they would receive a scorecard before the start of the treatment, but the first scorecard was followed by a phone call to the ACL where the indicators were explained and the ACLs could ask questions about the scorecard. The scorecards are also accompanied by an explanatory note showing how the numbers in the scorecard are calculated and a phone number to call to ask questions about the scorecard.

2.3.1 Additional intervention: List of peer performances

To test for peer effects, an addition was made to the scorecards for 77 randomly selected treatment offices in September 2019, a year after the first scorecards were sent out. The purpose was to test if there was an additional effect, beyond the effect of the scorecard, stemming from a bureaucrat's performance being observable to the bureaucrat's peers at the same position in the organizational hierarchy. For a randomly selected group of 77 offices within the offices already receiving the performance scorecards, a list of the percentile rankings of the two performance indicators for all 77 offices was added to the scorecard. Appendix Figure A5 shows an example of the first page of such a list. The main difference between receiving the typical scorecard and the scorecard with the list of performances was that for the offices that received the list of performances, their performance was observable not just to them and their supervisors but also to 76 of their fellow ACLs.

2.4 Randomization

Figure 2 provides a visual overview of the randomized interventions and data sources. The randomization was done in two waves. In August 2018, 112 land offices were using the e-governance system. In the first randomization wave, 56 of these offices were randomly chosen to receive the performance scorecards, while 56 were assigned to the control group.[7] In April 2019, 199 additional offices had started to use the e-governance system and a second randomization wave was carried out to increase the experiment's sample size. The second randomization wave extended the treatment to 99 new offices while 100 new offices were added to the control group.[8] The additional list of peer performances was added to the scorecard for 77 randomly selected offices receiving the scorecards in September of 2019. The scorecards were sent out until March 2020, when the outbreak of COVID-19 caused an end to the scorecards being sent out.

Both randomization waves were stratified by the number of applications processed within 45 working days in the two months preceding the randomization and the number of applications pending for more than 45 working days at the end of the month preceding the randomization. For the first randomization, another binary variable for being a land office where the e-governance system was fully implemented, meaning that no applications were conducted using the traditional paper-based method, was also used for stratification. In the second randomization, the total number of received applications was used as a stratification variable. The randomization of offices into receiving the peer performance list was done among the 155 offices receiving the scorecards using the same stratification variables as the second randomization wave. For more information about the randomizations see Appendix Section B.1.

2.5 Data

I use two main data sources, administrative data from the e-governance system and data from a survey conducted among applicants in the 112 land offices that were part of the first randomization wave. I use the administrative data to generate the performance scorecards as well as evaluating the effects of

[7]The first randomization was carried out by the author on 14 August 2018.
[8]The second randomization was carried out by the author on 10 April 2019.

the scorecards. Table 1 shows summary statistics for both data sets. This table contains all observations from both treatment and control offices that are used in the analysis. For a discussion of the balance of randomization, see Section 2.6.

2.5.1 Administrative data

The observations in the administrative data are at the application level. The data contains information about in which land office the application was made, the application start date, the date it was processed as well as the decision to accept or reject the application. The administrative data also contains information on how large the land plot for which the change is being made is.[9] The administrative data was downloaded from the e-governance system at the beginning of each month from August 2018 until October 2020.

For the main analysis, I use administrative data for applications from 13 August 2018, 1 month before the start of the experiment, until 20 January 2020. From 26 March 2020 and onwards the COVID-19 outbreak in Bangladesh substantially increased processing times for land records changes as measured by calendar days but also resulted in a large number of general holidays, increasing the difference between calendar days and working days. At this time, the scorecard intervention was also stopped. Therefore, I do not include applications made after 20 January 2020, 45 working days before the start of the general holiday caused by COVID-19, in the analysis. Ending the data at this point precludes the holiday from affecting one of the main outcomes, if the application was processed within the 45 working day time limit or not.

I impute the processing times for the 6% of applications that have not yet been processed. The imputed value is the mean of actual processing times that are larger than the number of working days the application that I am imputing the processing time for has been pending.[10] The data set in the main analysis contains 1,050,924 applications from all 311 offices. Appendix Section B.2.1 provides more information about the administrative data.

[9]The full administrative data set also contains more information about the applicants, but this data is not available for research purposes due to privacy concerns.

[10]This procedure is conservative in two ways. First, it reduces any effect on processing times generated by the scorecards since the same mean is used to impute values in both the treatment and control areas. Second, the mean used to impute processing times in this procedure likely underestimate the time it will take to process these applications on average since it is the mean of applications that *have already been processed*, which is likely to be less than the actual average time it will take to process all applications including those currently pending. Since the point estimate of the scorecards' effect on the share of applications being pending is a decrease of 0.9 percentage points, using these imputed values creates a conservative estimate of the effect of the scorecards on processing times.

2.5.2 Survey data

The survey data was collected in two rounds from applicants who applied in the 112 offices that were part of the first wave of randomization. The sample of applicants was created by placing surveyors outside land offices and interviewing all applicants entering the office for the purpose of a land record change application, regardless of what stage in the application process they were at. The surveyors stayed outside a specific office for at least two days and until they had completed at least 20 interviews. The follow-up interview was conducted by phone approximately three months after the initial interview. Surveyors were not informed about which offices had received the scorecards or if they were calling a respondent from a treatment or control office.

Out of 3,696 people approached, a total of 3,370 applicants were successfully interviewed in the first round interview outside of the land offices. Out of those interviewees, 3,018 were successfully re-interviewed in the follow-up phone interview, resulting in a total attrition rate of 18%. The estimated effect of the scorecards on the attrition rate was 3 percentage points and marginally statistically significant at the 10% level. However, in Appendix Section B.2.3 I show that this differential attrition is not sufficiently large to substantially affect the main findings from the survey data. More information about the survey data can be found in Appendix Section B.2.2.

The initial interview focused on the details of the application, the applicant's expectation for the application processing time, the applicant's willingness to pay for faster processing, as well as basic information about the applicant. The follow-up interview focused on the outcome of the application and the payments, above the official fee, that the applicant had made in relation to the application.

Data on bribe payments was collected using two different questions. The first question asked what the typical bribe payment is "for a normal person like yourself." If the respondent were willing to answer this question, the amount, whether zero or positive, was recorded as the variable *typical payment*. 63% of respondents provided an answer to this question and the average response was BTD 6,731 (~USD 80) or 1.5 months of the sample's average per capita household expenditure.[11] 73% of the responses were non-zero amounts. The second set of questions asked about each actual payment made by the applicant to any government official or agent assisting with the application. The outcome variable *reported payment* is the sum of the bribe amounts reported in each of these questions. This variable takes the value zero when no payments were reported. Since the most common response for respondents who were not

[11] Variables are winsorized at the 99th percentile and averages are calculated using observations weighted by the inverse of the number of observations in each office.

willing to talk about payments that they had made was to report no payment, as opposed to stating that they did not want to respond to the question, the average reported payment is likely an underestimate of the actual payments. The average reported payment was BDT 1,456 (~USD 17) and 27% of respondents provided a non-zero value. Among those reporting a non-zero amount the average amount was BDT 5,283 (~USD 63).

2.6 Balance of randomization

Appendix Table A1 shows a balance of randomization test for the two main outcome variables from the administrative data, the fraction of applications processed within 45 working days and the average processing time. The data used is restricted to applications made at least 45 working days before the start of the experiment. Applications that were not processed by the start of the experiment were assigned an imputed processing time, using the imputation procedure described in Section 2.5.1. There are no statistically significant differences between scorecard and control offices before the start of the experiment. This is expected given that the random treatment assignment.[12]

Appendix Table A2 shows that the scorecards did not affect the composition of applicants or applications in the survey data. This is not a traditional balance of randomization table, since the treatment may have affected which applicants decided to apply and what type of applications to make. However, I do not find any evidence for such changes in behavior. I find no statistically significant difference in the age or income of the applicants, or in the size or value of the land that the applications are for. Furthermore, there are no substantial differences between the stages that the applications are in at the time of the first interview. When using the regression specification from Equation 1 on this data, the effect of the scorecards is not significant at the 5% for any of the outcome variables, and significant at the 10% level only for land value.[13]

2.7 Additional intervention: Providing information to applicants

Together with the in-person survey, an intervention providing additional information to applicants was also carried out on randomly selected days in each office where the survey took place. The motivation

[12]Using the empirical strategy described in Section 3.1 on the data from before the start of the experiment also generates statistically insignificant estimates of the effect of the treatment on the outcome variables. Furthermore, an F-test of joint significance for the explanatory power of the outcome variables on the treatment variable cannot reject the null of no explanatory power (p-value: 0.69).

[13]F-tests of joint significance for the explanatory power of the outcome variables on the treatment variable cannot reject the null of no explanatory power (p-value: 0.73).

behind this intervention was to ensure applicants knew about the improvements in processing times. While it is likely that this information would eventually have spread, in the short-term, information about changes to bureaucrat behavior may not yet have disseminated. If the applicants are not aware of the improvements in processing times, the long-term effects on bribe payments may not yet have been realized. To speed-up the dissemination process, and potentially reach the long-term effect of the scorecards faster, the surveyors randomly provided information about increased processing speeds on half of the days that the in-person survey was conducted. The surveyors used an information pamphlet to inform applicants that the median processing time for all land offices had been substantially reduced over the past six months and that a new e-governance system had been installed. The information the surveyors provided was the same in both treatment and control offices. The scorecard intervention was not mentioned to applicants. Appendix Figure A6 shows an English translation of the information pamphlet. I will analyze the results of the intervention when testing the predictions of models connecting processing times and corruption in Section 5.

3 Empirical Strategy

3.1 Empirical strategy: Overall effects

To estimate the effects of the scorecards, I use the following regression specification:

$$Outcome_{ait} = \alpha + \beta Treatment_i + Strata_i + Month_t + \varepsilon_{ait} \quad (1)$$

Where $Outcome_{ait}$ is an outcome for application a, in land office i, made in calendar month t. $Strata_i$ are randomization strata fixed effects. Since no randomization strata overlap the two randomization waves, these fixed effects also control for randomization wave fixed effects. $Month_t$ are fixed effects for the month the application was made. In the survey data, all continuous variables are winsorized at the 99th percentile.[14] Standard errors are clustered at the land office level resulting in 311 clusters in the administrative data and 112 clusters in the survey data. Each observation is weighted by the inverse of the number of observations in land office i. Therefore, the estimated effect is the average effect of the scorecard on a land office, the level at which the treatment was assigned. The weighting also improves

[14] In the survey data, the application month variable is winsorized at November 2018, so that all application dates before November 2018 take the value of November 2018. A separate dummy variable controls for missing start date values.

the estimates' precision by making each cluster have equal weight in the analysis.[15]

3.2 Empirical strategy: Heterogeneous effects

To better understand the mechanisms behind the overall effects, I separate offices by their baseline performance and estimate the effect of the scorecards separately for offices performing above and below the median at baseline.[16] I calculate each office's baseline performance based on the average of the two percentile rankings at the time of the first scorecard. One ranking is based on the number of applications disposed within 45 working days, while the other is based on the number of applications pending for more than 45 working days. For offices in the treatment group, these are the actual rankings shown on the first scorecard, while for the control group, the rankings were not shown to the bureaucrats. I then separate all offices into *over-performers*, that were above the median average ranking at baseline, and *under-performers*, that were below the median average ranking at baseline.[17] Since the classification of offices only uses data from before the first scorecard was delivered, it is not affected by the treatment.

I use the following regression specification to estimate the effect of the scorecards on the two types of offices separately:

$$\begin{aligned} y_{ait} =& \alpha + \beta_1 Treatment_i \times Overperform_i + \beta_2 Treatment_i \times Underperform_i + \\ & \gamma Overperform_i + Stratum_i + Month_t + \varepsilon_{ait} \end{aligned} \tag{2}$$

Where β_1 is the estimated effect of the scorecards for offices over-performing at baseline, β_2 is the effect for offices under-performing at baseline, and γ is the difference between over-performing and under-performing offices in the control group.[18] As in the estimation of the overall effects, standard errors

[15]For a discussion of why weighting observations by the inverse of the number of observations in a cluster improves precision see: https://blogs.worldbank.org/impactevaluations/different-sized-baskets-fruit-how-unequally-sized-clusters-can-lead-your-power

[16]Heterogeneity in the effects of performance information provision between high and low performers has been recorded in several settings (e.g., Allcott, 2011; Dodge et al., 2018; Ashraf, 2019; Barrera-Osorio, Gonzalez, Lagos, and Deming, 2020). This was the only heterogeneity test based on office characteristics specified in the pre-analysis plan. The two other pre-specified tests for heterogeneity were based on the date of application and the application processing time. The estimates of heterogeneity in the effects along those dimensions are shown in Figure 4 and Appendix Table A3, respectively.

[17]I classify offices in the first randomization wave into over- and under-performers by comparing them to the median performance among these 112 offices at the time of their first scorecard (September 2018). For the offices in the second randomization wave, I compare them to the median performance of all 311 offices in the experiment at the time of their first scorecard (April 2019). This ensures that the over- and under-performer classification corresponds to if the content in the first scorecards was above or below the median of comparison groups at the time.

[18]To test the hypothesis that the treatment had the same effect on offices over-performing and under-performing at baseline, I use a similar regression but where the first treatment variable is not interacted with the dummy variable for if the office over-performed at baseline. I then test the hypothesis that the coefficient on the treatment variable interacted with with the dummy variable for if the office was under-performing at baseline is zero. This test's p-value is reported as "P-value sub-group diff." in the regression tables reporting the heterogeneous effects.

are clustered at the land office level and the regressions are weighted by the inverse of the number of observations in land office i.

3.3 Analysis of additional experiments and potential interactions

The two additional randomized interventions, the addition of peer performance lists and the information intervention to applicants, are not included in the main specification as these interventions are not the main treatments being evaluated. For the two main outcomes, delays and bribe payments, the full specifications, including the scorecard treatment, the additional randomization, and the interaction, can be found in Tables 3 and A4. These tables show that neither of the two additional experiments have substantial interactions with the scorecard treatments, validating the approach to analyze the scorecard treatment separately as outlined in Equations 1 and 2.

4 Results: Effects on Processing Times, Bribes and Visits by Applicants

This Section shows the estimates of the effects of the scorecards on processing times, visits to land offices made by applicants, and bribes. Appendix Section C.3 investigates potential unintended consequences of the scorecards on bureaucrats' behavior and does not find evidence for any large unintended consequences.

4.1 Effect on processing times

Table 2 shows that the scorecards increased the applications processed within the government time limit and improved processing times overall. Each column presents the result of a regression using the specification in Equation 1. Column (1) shows the estimated effect of the scorecards on a binary variable indicating if the application was processed within the 45 working day time limit or not. The scorecards increased the fraction of applications processed within the 45 working day limit by 6 percentage points or, equivalently, 11%. Column (2) shows the estimated effect on the Inverse Hyperbolic Sine (IHS) transformation of the number of working days it took to process the application.[19] Column (2) estimates that the scorecards reduced the processing time by 13%.[20] In the data, 6% of the applications are not

[19]The IHS transformation is used instead of the natural logarithm since 0.3% of the applications were processed on the same day as they were made and therefore have a processing time of zero working days. The results are virtually identical when dropping the applications taking zero days to process and using the natural logarithm transformation.

[20]The exact effect is 13 IHS points, which are approximately equivalent to log points. A 13 log point decrease is equivalent to a 12% decrease, but for simplicity, I will describe IHS points changes as percentage changes throughout the paper. Appendix Table A6 shows that the result is similar when dropping the observations with processing times of zero working days and

yet processed, and for the analysis in Column (2) I have assigned imputed processing times for these applications, using the imputation procedure described in Section 2.5.1. Appendix Table A5 shows that the results are robust to different imputation techniques. In Appendix Table A6 I test the robustness of the result to using different functional form assumptions for the relationship between the scorecards and processing times.

For Column (3), I create an Inverse Covariance Weighted (ICW) index of the two outcomes used in Columns (1) and (2).[21] The estimated effect of the scorecards on the ICW index is 0.13 standard deviations and statistically significant. In Appendix Table A7 I test the robustness of this result with various alternative specifications. All alternative specification estimates are of the same sign and similar magnitude as the main estimate, but some of them are not statistically significant. Appendix Table A8 shows the effects, estimated at the office by month level, on the number of applications processed within 45 working days, the number of applications pending beyond 45 working days as well as those figures corresponding percentile rankings. The point estimates suggest that the scorecards improved all four of these outcome variables but the effects are not statistically significant.

4.1.1 Effect over time

Figures 3 and 4 show that there is no pattern of the effect declining over time, although the size of the effect varies between different time periods. Figure 3 shows the fraction of applications processed within the 45 working day limit over time for the treatment and control group separately. The first dashed vertical line indicates the date 45 working days before first scorecards. The second dashed vertical line indicates the date of the first scorecards. Applications made between the first and second vertical lines may have been affected by the scorecards if they were not processed before the first scorecard was sent out. Starting for applications made a few days before the first scorecards, we see a divergence between the treatment and control group. The treatment group increased the fraction of applications that were processed within the 45 working days time limit, relative to the control group. With a few short exceptions, the treatment offices continue to have a higher fraction of applications processed within the time limit relative to the control offices until the end of the experiment. The data for the offices in the second

using the natural logarithm transformation.

[21] The ICW matrix follows the algorithm suggested by Anderson (2008) and is designed to summarize several outcome variables into one index that, for the control group, has a mean of zero and a standard deviation of one. Since there are only two outcome variables in Table 2, the ICW index is equivalent to summing the standard deviations away from the control group mean of the two variables and rescaling the index to have a standard deviation of one in the control group. However, in tables with more than two outcome variables, the components are weighted differently to maximize information captured by the ICW index.

randomization wave ends earlier relative to the start of the experiment. The third vertical dashed line marks where the data from the second randomization wave ends. To the right of this line, the graph only contains data from the offices in the first randomization wave. Appendix Figure A7 shows the time lines for the two randomization waves separately.

Figure 4 shows the results of applying the regression specification from Equation 1 to applications made in the first, second, and last third of the experiment period. The outcome variable is the ICW Index from Column (3) of Table 2. When I split up the sample, the estimates lose some precision, but it is clear from the graph that there is no pattern of a continuous decline of the effect over time.

4.1.2 Effect on the distribution of processing times

Figure 5 shows two overlaid histograms, one for the distribution of processing times in the treatment group and one for the distribution in the control group. The figure only includes applications that have already been processed and processing times are top coded at 200 working days. In the treatment offices, more applications were processed within the 45 working day time limit. The effect is relatively evenly spread over the whole span from 0 to 45 working days, with only a minor bunching just before the 45 working day limit. This is to be expected given that the process to approve an application is relatively long and depends on several individuals, as described in Section 2.1. This means that even if the ACL targets a 45 working day processing time, there will be a considerable spread around this target. Because of this, the ACLs may target a processing time lower than 45 working days. The figure also shows that the processing times that are reduced in frequency by the scorecards are in the whole span from 55 working days and up. This is also reasonable given that the scorecards emphasized both processing applications within the 45 working day limit and reducing the number of applications pending beyond 45 working days. Overall the spread of the effect in the distribution of processing times alleviates the concern that ACLs are "gaming" the scorecards by only speeding up the processing of applications that would otherwise have been processed within a few working days outside of the time limit.

4.2 Mechanisms for the effect on processing times

The scorecards increase the information the bureaucrats and the bureaucrats' supervisors have about the performance of the bureaucrat. This could improve performance through two main channels. First, the supervisors may improve the incentive structures the bureaucrat is facing by facilitating better promotions and more attractive postings for those bureaucrats with good scorecards, or more generally, bureaucrats with a good overall reputation of which the scorecards are a part. This is an example of the widely studied mechanism of increased information enabling better contracts that improve output (Holmström, 1979). It is also possible that bureaucrats care about their supervisors receiving information about them for other reasons, such as the shaming effect of having a negative performance being shown to a superior.

Second, bureaucrats may change their behavior due to receiving the scorecards themselves. For bureaucrats, receiving information about their delays each month may increase this information's salience, causing it to be more important for their personal sense of shame or pride in their work.[22] Since the scorecards were sent to both bureaucrats and their supervisors, I cannot separately estimate the importance of these two mechanisms and I refer to them collectively as *reputational concerns.*

In addition to the two mechanisms above, it is also possible that information flows between bureaucrats at the same level in the organizational hierarchy create an additional incentive for improved performance through peer effects (Mas and Moretti, 2009; Bandiera, Barankay, and Rasul, 2010; Cornelissen et al., 2017).[23] I estimate the magnitude of such a peer effect, above and beyond the effect of the scorecard, by sending information about other offices' performance within a randomly selected sub-group of the offices receiving scorecards, as described in Section 2.3.1.

Table 3 shows the effect of the peer performance list intervention on processing times. Sharing the performance information of a bureaucrat with other bureaucrats does not meaningfully improve processing times beyond the effect of the performance scorecards. Column (1) of Table 3 shows that the estimated effect on the fraction of applications processed within the 45 working day time limit is positive but close to zero. Column (2) shows that the effect on overall processing times is negative but also close to zero.[24]

[22] Effects from simply being informed of one's own performance have been found for energy conservation (Allcott, 2011). On the other hand, the effects of such information provision in private organizations have been mixed, with several papers showing that even the direction of the effect depends on the specific circumstances (Blader, Gartenberg, and Prat, 2020; Ashraf, 2019).

[23] In addition to the context of job performance, effects of sharing information about behavior to others have shown to improve socially desirable behaviors such as voting (Gerber, Green, and Larimer, 2008) and paying taxes (Bø, Slemrod, and Thoresen, 2015; Perez-Truglia and Troiano, 2018).

[24] Columns (3) and (4) of Table 3 use the full data set and estimate the effect of the scorecard and the peer performance list simultaneously. This is done using a dummy variable for the peer performance list treatment that takes the value of one for applications made in offices receiving the peer performance lists, made later than one calendar month before the first performance list was sent out. When estimating the effects of the scorecards without the effect of the performance list, the point estimates are similar to the effect in the main estimate but only statistically significant at the 10% level. This shows that the effect of the scorecards is not driven by the inclusion of the peer performance list.

4.3 Effect on visits to land office and time spent by applicants

In Table 4, I use survey data to show that the scorecard reduced the number of visits to land offices by the applicants as well as the total hours spent on making these visits. Column (1) shows that the scorecards reduced the number of visits by 1.0 visits, or 12%. Column (2) estimates that the scorecards decreased the total number of hours spent on these visits by 1.6 hours, or 7%, but this effect is not statistically significant. Column (3) estimates the effect on an ICW index of these two outcome variables showing that the effect is not statistically significant for a combination of the two variables. Appendix Section C.5 shows that the scorecards did not improve the stated satisfaction with the application process among applicants.

4.4 Effect on bribe payments

Table 5 shows that the scorecards did not lead to a decrease in bribe payments. Instead, the estimated effect on bribes is positive, although this increase is not statistically significant. As described in Section 2.5.2, data on bribe payments was collected using two separate survey questions. The first question asked about the typical bribe payment "for a normal person, like yourself." When this measure is used, the column is marked as "typical." The second set of questions asked about each payment made by the applicant. When this measure is used, the column is marked "reported."

Columns (1) and (2) of Table 5 show the effect on the amount of bribes paid. Column (1) shows that the effect on the perceived typical payment was BDT 1,046 (USD 12), a 17% increase, statistically significant at the 10% level. Column (2) estimates that the scorecards increased reported bribe payments by BDT 265, a 21% increase, but the result is not statistically significant. Columns (3) and (4) show that there is no effect on the propensity to report a non-zero bribe. This can be interpreted as the scorecards having no effect on the extensive margin of bribe payments. Another interpretation is that the intervention did not affect applicants' willingness to talk about bribe payments in the survey. In Columns (5) and (6), the sample is restricted to those who reported non-zero bribe payments. Bribe payments increased by 19% for typical payments and 23% for reported payments, with both effects being statistically significant. Again these effects have two interpretations. Either the scorecards only affected the intensive margin of bribe payments, or the scorecards increased bribe payments for at least those applicants who were willing to describe what bribes they paid but potentially also for other applicants. The estimated effects for a range of alternative specifications for the main estimate in Columns (1) and (2) of Table 5 are shown in Panel A of Appendix Table A9. All alternative specification estimates are qualitatively similar, but some are of slightly larger magnitude and, therefore, statistically significant.

4.5 Heterogeneity of results by office performance at baseline

I use the empirical strategy described in Section 3.2 to understand if there are differences in the effect of the scorecard between offices over- and under-performing at baseline.

4.5.1 Heterogeneity in effects on processing times, visits, and time spent by applicants

Table 6 shows that the effect of the scorecard on processing times is driven by offices that were under-performing at baseline. Column (1) of Table 6 shows that for offices that were over-performing at baseline, the effect on the fraction of applications processed within the 45 working day limit was just 0.8 percentage points. For offices that were under-performing at baseline, the effect was 12 percentage points, equivalent to a 30% increase. Column (2) shows that for offices over-performing at baseline, the effect of the scorecard on the total processing time is a decrease of approximately 3%. For offices that were under-performing at baseline, the effect was a decrease of 23%.

The effects in the survey data are less precisely estimated but also show that it is offices underperforming at baseline driving the effect. Column (3) of Table 6 shows that for offices over-performing at baseline, the number of visits per applicant was reduced by 0.7 visits, while for offices under-performing at baseline, the effect was a decrease of 1.2 visits, equivalent to a 12% decrease. Column (4) shows that for offices over-performing at baseline, hours spent on the application by applicants increased by 0.4 hours while in offices under-performing at baseline the effect was a decline of 3.0 hours, equivalent to an 11% decrease.

Overall it is clear that the improvements that the scorecards led to were almost entirely driven by offices that were under-performing at baseline. Appendix Table A10 shows that this result is robust to other measures of baseline performance. Panel B of Appendix Table A7 shows that this result is robust to alternative regression specifications. There are several reasons for why over-performing offices may respond more to the scorecards. For example, negative performance feedback may create a stronger desire to improve one's performance for subsequent scorecards. However, it may also be the case that poorly performing offices have a larger scope for improvement since there is more "low-hanging fruit" in terms of increasing efficiency.

4.5.2 Heterogeneity in the effect on bribe payments

Table 7 shows that the positive effect on bribe payments is entirely driven by the offices that were over-performing at the start of the experiment. Column (1) shows that the effect of the scorecard on estimated typical bribe payments among offices over-performing at baseline was an increase of BDT 2,280 or 43% and statistically significant. Column (2) shows that the reported payments among offices over-performing at baseline increased by BDT 638 or 70%, and is statistically significant. The effect on offices that were under-performing at baseline is close to zero and not statistically significant. Appendix Table A10 shows that this result is robust to other measures of baseline performance. Panel B of Appendix Table A9 shows that this result is robust to alternative regression specifications.

This result is surprising, given that over-performing offices did not change their behavior in terms of processing times. I will discuss this result at length in sections 5 and 6.

4.6 External validity, potential biases from surveying, and unintended consequences[25]

One advantage of the design of the experiment is that it was conducted at a large scale, with more than half of Bangladesh's land offices taking part in the experiment. The large scale of the experiment makes it plausible that the results are externally valid within Bangladesh (Muralidharan and Niehaus, 2017). Offices took part in the experiment if they had the e-governance system installed. Therefore, the main concern for the external validity of the result within Bangladesh is that offices that had the e-governance system installed earlier had a larger effect than offices where the e-governance system was installed later. Appendix Section C.1 shows evidence that the effect of the scorecards on processing times was only slightly larger in the land offices that had the e-governance system installed earlier. Furthermore, using a linear prediction, the effect is predicted to be positive for all offices in Bangladesh where the e-governance system is installed.

While it is unlikely, it is possible that the survey and information intervention affected the overall effect of the scorecards. In Appendix Section C.2 I restrict the sample to applications made before the survey took place and application made in offices where there was no survey and show that there is no evidence that the survey or information intervention are drivers of the estimated effect of the scorecards on processing times.

A common problem of quantitative performance measures is that they often lead to gaming of the quantitative measures or other unintended consequences (e.g., Banerjee, Duflo, and Glennerster, 2008;

[25] Appendix Section C discusses potential biases, external validity, and unintended consequences in detail.

Rasul and Rogger, 2018). In Appendix Section C.3 I test for three such potential unintended consequences that could have improved the scorecards without increasing the real service delivery speed for applicants. First, if bureaucrats allow fewer applicants to start applications, then this may improve their scorecards, provided that the lower number of applications help them process a larger share of the applications within the time limit. Second, if bureaucrats allowed applications selectively such that the average application was easier to process within the time limit, then this could have improved their scorecards. Finally, the scorecards may lead to bureaucrats making worse decisions regarding accepting or rejecting applications. Reassuringly, I do not find any evidence for any of these unintended consequences.

5 Implications for Theories of Processing Times and Bribes

In this Section I will show how the experimental results are inconsistent with several common models of how bribes are related to delays in public service delivery, or more generally red tape.[26] There are several theoretical reasons for why bribes may be causally related to delays. Some of these models predict a positive causal relationship, while others predict a negative relationship. For example, fast processing times may increase the applicants' willingness to pay for the public service and enable bureaucrats to extract more bribes. Bribes may also provide a piece rate incentive for bureaucrats to process more applications and cause bureaucrats to process applications faster. Conversely, long processing time for those paying small or no bribes may enable bureaucrats to extract more bribes from applicants willing to pay to get their application processed fast. These causal relationships exist both in models where corruption is efficiency-enhancing (e.g., Leff, 1964; Huntington, 1968; Lui, 1985), as well as in models where corruption is the original cause of the slow service delivery (Myrdal, 1968; Rose-Ackerman, 1978; Kaufmann and Wei, 1999).

It is important to understand which, if any, of these relationships are major determinants of bribes and processing times since some of the models have opposing policy implications. If slow service delivery causes corruption, then expanding the processing capacity of the bureaucracy through more staff, better technologies, or better management, may not only improve processing times but also reduce corruption. But if bribery was to be rooted out, without addressing the underlying capacity constraints, this might lead to a worse situation for applications if, for example, bureaucrat had less of an incentive to

[26] I use the term delays, but the theories are equally applicable to other forms of red tape, such as the need for multiple visits to government offices or an excessive amount of paperwork to be filled out.

process their applications or people who urgently needed a service could not pay to get it faster. On the other hand, if corruption is the underlying cause for slow service delivery due to intentional delays by bureaucrats for the purpose of extracting more bribes, then providing the bureaucracy with more staff or better technology would not lead to any improvement in processing times, let alone decrease corruption. The most important policy priority should then instead be to eliminate corruption to remove the incentives for bureaucrats to intentionally delay corruption.

The model the scorecard experiment was originally designed to test was a model of a direct causal relationship where the presence of corruption led to slower processing times but where faster processing times could reduce corruption. The model was similar to monopolistic price discrimination models, and in the model bureaucrats use delays strategically to maximize the total amount of bribes in the same way a monopolist would strategically decrease the quality of some goods to maximize profits (Mussa and Rosen, 1978; Maskin and Riley, 1984). It assumes that applicants have different willingness to pay to avoid delays, but that bureaucrats cannot perfectly observe the willingness to pay of each applicant. Therefore, they intentionally delay applications from applicants only paying low bribes in order to extract more bribes from applicants with a high willingness to pay to avoid delays. The model predicts that an improvement in processing times, such as the improvement the scorecards created, should decrease bribe payments among applicants getting their applications processed the fastest. See Appendix A.2 for a more detailed description and an explicit test rejecting this model in this context.

A different type of models, are models where the government officials could extract more bribes if they wanted to, but choose not to do so because there is a trade-off between taking bribes and some other objective of the government official. This trade-off could be between taking bribes and the risk of getting caught (Becker and Stigler, 1974; Olken, 2007; Niehaus and Sukhtankar, 2013a), but it could also be a trade-off between bribes and altruistic or social motivations for not taking bribes. In Section 6, I develop a specific such model where taking bribes hurts bureaucrats' utility through bribes negative effect on reputation but where this can be compensated for by better visible job performance in terms of processing times. I then derive predictions and test these against the results of my experiment.

5.1 Do faster processing times decrease bribe payments?

The results in Section 4 are not consistent with theories of a causal relationship between faster average processing times and lower bribe payments. While the scorecards did reduce processing times, it did not reduce bribes, as shown in Tables 2 and 5, respectively. This is true even for the offices that were under-performing at baseline and improved their processing time the most, as shown in Tables 6 and 7.

One potential reason for the lack of effect from the scorecards on bribe payments could be that the information about the improvement in processing times had not yet disseminated among applicants. There are two reasons why this is not plausible. First, the scorecards did decrease expected processing times. Column (1) of Appendix Table A11 shows the effect of the scorecards on the expected total processing time at the time of the first survey interview. The scorecards reduce expected processing times by 9%, similar in magnitude to the effect on actual processing times and statistically significant. Second, to further rule out that the lack of information about the improved processing times limits the effect of the scorecards on bribes, I use the information treatment that was designed to inform applicants about improvements in the processing times, as described in Section 2.7. Column (2) of Appendix Table A11 shows that the point estimate for the effect of the information intervention on expected processing times is a reduction of 4% but that the estimate is not statistically significant. Taken together, these results suggest that applicants are aware of the current processing times in their sub-district land office and that providing them with more information does not substantially change their expectations. Appendix Section C.4 shows that the information treatment did not affect bribes, neither by itself nor in combination with the scorecards.

5.2 Implications for other theories

Given the positive effect of the scorecards on bribes, is it possible that faster average processing times lead to higher bribes? Tables 6 and 7 show that for offices under-performing at baseline scorecards improved processing times the most but did not change bribe payments. This is inconsistent with any model where average processing times has a causal effect on bribe payments. Furthermore, for the offices over-performing at baseline, the scorecards increased bribe payments without changing the processing times, which is inconsistent with models of a causal effect of bribes on processing times.

The increase in bribes among the offices that were over-performing at baseline is also inconsistent with models where it is an applicants' outside option or ability to pay that determine the bribe levels (Svensson, 2003; Niehaus and Sukhtankar, 2013b). If bribe levels change as a result of a positive scorecard sent to the government official responsible for the service for which the bribe is paid, without any observable change in service quality. The bribe level cannot be fully determined by the applicants' outside option or ability to pay. This result is most likely dependent on the structure of the interaction in which the bribe is paid. In this context, the land office is the only institution that can make the required land record change and there are no close substitutes to this service. Therefore, the bribe level is expected to be determined mainly by other factors. If there had been competition for applicants between land offices, or a close alternative to a land record change, it is plausible that these outside options (or "exit" options) would have been more important in determining the bribe level (Svensson, 2003).

6 Model of Bureaucrat Behavior

In this section, I will provide an overview of the model I propose to explain the results of the experiment. Appendix Section A.1 provides a formal presentation of the model.

6.1 Model set-up

In the model, bureaucrats get utility from a reputational concerns term which is a function of visible job performance in terms of delays and bribe money.[27] Bureaucrats get disutility from effort, but effort is needed to avoid delays, which decrease the reputational concerns term. Reputational concerns has decreasing marginal utility, and so does bribe money, while effort has increasing marginal disutility. Bribes and delays both reduce the reputational concerns term since if a bureaucrat consistently asks for high bribes and do not process applications on time, a negative reputation about the bureaucrat is built and becomes visible to others. The visibility of delays increases with the scorecards, making delays more important for the bureaucrats' reputational concerns.

Bureaucrats differ only in the extent to which they care about their reputational concerns. This could be because of differences in discounting future career prospects, differences in the valuation of social status from holding a high-level civil service position, or differences in intrinsic motivation to "do a good job." What is important about these differences for the model is that they create the difference between over- and under-performing bureaucrats.[28] This assumption is also consistent with the observation that over-performing bureaucrats collect less bribes than under-performing bureaucrats in the control group. This would not be the case if ability is what made over-performing bureaucrats better

[27]The reputational concerns term represents reasons for why the bureaucrats cares about what others, especially their supervisors, think of them. This could be for material reasons, such as career progression, social reasons, such as maintaining a good social standing with others in the bureaucracy, or psychological reasons such as the negative feelings of pride (or shame) stemming from knowing that someone else knows about one's good (or bad) performance. The term also encapsulates psychological reasons that are internal to the bureaucrats, such as the negative feelings stemming from failing to perform one's duty or breaking an internalized social norm of performing at least as well as one's peers. I.e., the effect of the scorecards on the reputational concerns term captures both possible mechanisms described in Section 4.2.

[28]Ashraf, Bandiera, Davenport, and Lee (2020) show that differences in the motivations of public servants is important for public service delivery.

than under-performing bureaucrats.[29]

In the model, the applicants simply pay the bribe amount that the bureaucrats are demanding. While this is clearly an abstraction from reality, Figure 1 shows that the average stated value of a record of rights for applicants is substantially higher than the values of bribes paid. Even the largest estimate for the average bribe, the estimate of a typical payment, is just 0.1% of the average estimated value of the record of rights. This difference between the applicant valuation and the amount paid suggests that the applicants' willingness to pay for the service is not an important determinant of the bribe value. Instead, what determines the amount of bribes that the bureaucrats extract in the model is the trade-off between bribe money and reputational concerns.

The model also abstracts away from bribes that increase the speed of processing and the value that fast processing-times have to the applicants. While this assumption is a simplification, Figure 1 shows that the average value an applicant put, even on the fastest reasonable processing time, is just 33% of the average estimated value of a typical bribe payment. This suggests that most of the bribes are not paid for increasing the speed of processing.

Finally, the model assumes that bureaucrats cannot buy reputation using money. This abstracts away from situations where applicants use bribe money to pay supervisors for promotions, but the results would be the same if bureaucrats would pay for the position as ACL in the first place but that they then cannot bribe their way to future career advancement or high social standing in the bureaucracy.[30]

6.2 Model Predictions

The theoretical model has two main testable predictions. In what follows, I describe these predictions, the intuition behind them, and how I test them empirically. Appendix Section A.1 provides the formal model as well as the derivations and formal statements of the predictions.

6.2.1 Effects of scorecard on delays

The first set of predictions relates to the effect of scorecards on delays. The scorecards have two different effects on delays, a *substitution effect* and an *income effect*. These effects are analogous to the substitution and income effects from a wage increase in a standard labor supply model. The substitution effect leads

[29] Table 7 shows that in the control group, offices under-performing at baseline also extract substantially larger bribes. This would not be the relationship if the differences in processing times were driven by a bureaucrat characteristic uncorrelated with bribe payments, such as ability. However, this relationship should be interpreted as an association, as the causal effect of the bureaucrat type on bribes is not identified by the experiment.

[30] Weaver (2020) analyses the effects of such bribes in the allocation of job applicants to positions in public service delivery.

to a decrease in delays for all bureaucrats. This is because the scorecards increase the importance of delays for bureaucrats' reputational concerns. Therefore, the marginal effect from decreasing delays increases and bureaucrats provide more effort to avoid delays.

The income effect from the scorecards on delays is positive for over-performing bureaucrats and negative for under-performing bureaucrats. For over-performing bureaucrats, the scorecards increase their reputation by making the already positive performance more visible. Since reputation has decreasing marginal utility, this decreases the marginal utility effect from changes to their reputation and reduces the optimal amount of effort they provide to avoid delays. Therefore, the model does not have a prediction for the effect of the scorecards on delays among over-performing bureaucrats, the direction of the effect depend on if the substitution or income effects is stronger. For under-performing bureaucrats, the scorecards make the negative performance more visible and make their reputation worse. This increases the marginal utility from reputation and hence increases the optimal amount of effort that bureaucrats provide to avoid delays. Hence, for these bureaucrats the substitution effect and income effect are in the same direction and the model predicts that the scorecards will reduce delays among under-performing bureaucrats.

Prediction 1:	Scorecards improve processing times for bureaucrats under-performing at baseline
Inconclusive:	Ambiguous direction of the effect on processing times for bureaucrats over-performing at baseline

These predictions are tested directly in Table 6, described in Section 4.5. Consistent with Prediction 1, the scorecards improve processing times for offices under-performing at baseline leading to fewer delays and shorter average processing times. The effect for over-performing offices is substantially smaller than the effect for under-performers and the difference in the two effects on the ICW index is marginally statistically significant.

6.2.2 Effects of scorecard on bribes

The second prediction relates to the effect of scorecards on bribe amounts extracted from applicants and the levels of these amounts in the control group. In the model, the bureaucrats could increase bribes by simply asking applicants for more money to approve their applications. Bureaucrats do not extract more bribes because of their reputational concerns. Therefore, the marginal utility from reputation is an important determining factor for bribe payments. When the scorecards improve over-performing bureaucrats' reputation, the marginal negative effect bribes have on utility through the reputational concerns declines. This leads to an increase in bribes taken by over-performing bureaucrats when they receive the scorecards.[31]

For under-performing bureaucrats, the decrease in the reputational concerns term leads to an increase in marginal marginal disutility from bribes coming through the reputation channel. This could lead to a decrease in bribes, but since effort increases in response to the scorecards, the overall effect on reputation could be positive or negative. Since the effect on bribes for this group is ambiguous, the model does not have a prediction for the effect of the scorecards on bribes for under-performing bureaucrats.

Prediction 2:	Scorecards increase bribes for bureaucrats over-performing at baseline
Inconclusive:	Ambiguous direction for the effect on bureaucrats under-performing at baseline

These predictions are tested directly in Table 7, described in Section 4.5.2. The scorecards increase bribes paid in offices over-performing at baseline. The effect of the scorecards on bribes is close to zero for office under-performing at baseline.

6.3 Potential general equilibrium effects within the civil service

In the predictions described above, I do not allow for the scorecards to change the benchmark performance that bureaucrats are compared against. In the context of the experiment, this does not qualitatively alter the predictions since half of the bureaucrats creating the benchmark do not receive the scorecards. However, if the scorecards were to be scaled-up to all bureaucrats, there would be a larger effect on the benchmark performance. This would shift the whole distribution of performance percentiles down and thereby have an income effect on all bureaucrats. The prediction from the model is that this income effect would induce more effort and smaller bribe payments than the partial experimental roll-out of the scorecards.

[31] This effect is dependent on that the income effect on delays is not so strong that it dominates the substitution effect and mutes any positive effect on the reputational concerns coming from the increased visibility of the already good performance. We see that this is not the case in Table 6, where the overall performance of bureaucrats over-performing at baseline is marginally positive, suggesting that the substitution effect marginally dominates the income effect and hence the scorecards lead to an increase in the reputation of over-performing bureaucrats by increasing the visibility of their positive performance.

6.4 Alternative explanations for the effects of scorecards

6.4.1 Increased marginal costs of bureaucrats' time or increased willingness to pay among applicants

One potential explanation for the scorecards resulting in increased bribe payments is that scorecards increase the marginal value of the bureaucrats' time. If bribe payments are made so that the bureaucrats spend more time on an application, the marginal value of the bureaucrats' time could be an important determinant of the bribe amount. If the scorecards increase the overall amount of time that bureaucrats are working, it is also likely that the marginal value of their time increased. Hence, it is possible for scorecards to have increased bribes through this mechanism. Another alternative explanation is that faster processing times lead applicants to be willing to pay more to get their land record change.

However, both of these explanations are inconsistent with the result that in the offices where the changes in processing times were the largest, bribe payments did not change. Instead, it was in the offices where changes in processing times were small that bribe payments increased. If it was an increase in the willingness to pay by applicants or an increase in bureaucrats' effort that lead to the increase in bribes, the increase would have taken place in the offices under-performing at baseline, because these were the offices where the scorecards improved processing times. Therefore, it is unlikely that either of these mechanisms is a substantial reason for the increase in bribe payments.

6.4.2 Transfers of over-performing bureaucrats

An alternative explanation, that is consistent with the heterogeneity in the effects on delays and bribes, is that over-performing bureaucrats get transferred due to receiving positive scorecards and that they are replaced by average performing bureaucrats. If the average performing bureaucrats both have slower processing times and collect more bribes, we expect that bribe payments would increase in offices over-performing at baseline. Processing times may not change as the incentive effects of the scorecards may cancel out the effect of high quality bureaucrats being replaced by lower quality bureaucrats.

However, this explanation is refuted by the data on bureaucrat transfers. Appendix Table A12 shows that the scorecards did not affect bureaucrats' transfers. Column (1) shows the overall effect on the probability of being transferred, Column (2) shows the heterogeneity in the effect by offices over-performing and under-performing at baseline. Columns (3) and (4) show the overall and heterogeneous effects on the duration of the posting for the first bureaucrat after the start of the experiment, including postings that started before the experiment. Columns (5) and (6) show the overall and heterogeneous effects on not having any ACL assigned to the office. All of the effects are close to zero and not statistically significant.

6.4.3 Over-performing bureaucrats using scorecards in negotiations over bribes with applicants

Another alternative explanation is that positive scorecards help bureaucrats prove to applicants that they have the ability to process applications quickly. This could then allow the bureaucrats receiving positive scorecards to charge higher bribes while it would not affect the bribes in offices reviving negative scorecards since these would not be shown to applicants.

There are three reasons why this explanation is implausible. First, the coefficients on "Overperform baseline" in Columns (4) and (5) of Appendix Table A11, show that the expected processing times are 14% lower in the over-performing offices not receiving the scorecards, suggesting that the applicants are already aware of the faster processing times in these offices. Furthermore, in Column (4) the point estimate for how the scorecards effect on applicants' expectations in over-performing offices is a 6% decrease, similar to the point estimate of a 3% decrease for the actual improvement of the processing times in these offices, as shown in Column (2) of Table 4. If the scorecards helped bureaucrats change applicants' expectations, the effect on the expectations should be larger than the effect on the actual processing times. Second, although I cannot rigorously rule out that no one in the land offices showed the scorecards to applicants, in none of the qualitative interviews done with ALCs and applicants was it even mentioned that the scorecards were shown to applicants and when directly asked, the applicants said they were not aware of the performance scorecards. Third, the information intervention tried to accomplish the effect that a bureaucrat could achieve by showing the scorecard to an applicant. Column (2) in Appendix Table A11 shows that the point estimate of the effect of the information intervention is just a 4% improvement in the expected processing time, suggesting that it is difficult to move applicants priors through simple information interventions. Furthermore, Appendix Table A4 shows that the information intervention did not increase bribes.

7 Conclusion

I have shown that information flows about individual government bureaucrats performance within a bureaucracy can improve the performance of these bureaucrats, even in the absence of explicit performance incentives. The results from the experiment show that these effects can happen rapidly and persist over at least 16 months. One plausible mechanism for this effect is that the bureaucrats care about the reputation they have among their supervisors. A second potential mechanism is that being measured and compared to your peers increases the salience of the performance to the bureaucrats themselves and generates a sense of shame or pride that create an additional motivation to perform well.

One way to assess the value of the improved processing times is to multiply the applicants' average stated valuation of having their application processed one day faster with the reduction in the total number of processing time days due to the scorecards.[32] For the 155 offices receiving the treatment, this gives a value of approximately USD 9.7 million per year.[33] This value should be interpreted carefully since it relies heavily on the stated value of faster processing to the applicants. However, the number is more than two orders of magnitude larger than the implementation costs of the scorecards, which were approximately USD 20,000 per year, even when including the author's time and set-up costs. However, the value of the experimental intervention becomes less clear when taking into account the effects on bribes. Multiplying the effect of the scorecards on reported payments with the number of applications in the treatment area results in an estimate of the effect on total bribes paid of 1.9 million per year. If the effect on the estimated typical payment is used instead, the total increase is USD 7.6 million per year.

Except for the increase in bribe payments, I do not find any evidence for unintended consequences or gaming of the scorecard's quantitative performance indicators. It is possible that monitoring or information flows that are not directly tied to explicit incentives are less likely to have the unintended consequences that are common for explicit incentive structures. One reason for this is that the receivers of the information can interpret the information flexibly. If bureaucrats engaged in observable behavior leading to unintended consequences, it would be possible for the supervisors to take this into account when interpreting the information on the scorecards. Furthermore, the scorecards were well received by most supervisors and there was no substantial backlash among bureaucrats. This points to another difference from explicit performance incentives. Since improved information flows do not reduce the discretionary power of supervisors, it is possible that they are less likely to be opposed by important actors in the organization and, therefore, poorly implemented (Banerjee et al., 2020).

The results have several policy implications. First, the result highlights that there exists untapped potential in data generated by e-governance systems. As more and more public services are delivered

[32] I calculate the value of having the application being processed one day faster using the following formula: $\frac{\text{Value of processing in 7 days}}{\text{Expected processing time from survey date}-7}$. All the information comes from the in-person survey made before the application was actually processed.

[33] The number of applications per year is estimated by taking the number of applications in the last six months of 2019 when all offices had the e-governance system installed and multiplying by 2.

using e-governance systems, the cost of monitoring and evaluating civil servants' performance has drastically decreased. The results of the experiment show that using the data generated by e-governance systems for monitoring and evaluation has significant potential to improve bureaucratic efficiency.

Second, the differential effects of the scorecards on under-performing and over-performing offices suggest that it is especially important to improve information flows for under-performing bureaucrats. Regardless of the mechanism for this result, it implies that the type of recognition systems that are common for bureaucrats in low- and middle-income countries, where outstanding performances are recognized without addressing inadequate performances, are ineffective. This is because providing positive feedback has a negative effect stemming from the improved reputation that the positive performance information generates. Instead, it is more important to make sure negative feedback is provided to under-performing civil servants. Positive feedback might still have an overall positive effect since it may motivate under-performing bureaucrats who want to receive better feedback, but the positive feedback is likely less effective than the negative feedback and can, in some cases, even be counter-productive.

Finally, the model points out a more general problem when using reputational concerns to incentivize a socially desirable behavior by an agent. Any reform or intervention that increases the reputation of some agents may also have a negative spill-over on other behaviors where reputation is a motivating factor. This is an especially important insight for government bureaucracies, where compressed wage structures and secure employment of civil servants often make reputational concerns more important motivators than in other organizations.

References

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics 95*(9-10), 1082–1095.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484), 1481–1495.

Ashraf, A. (2019). Do performance ranks increase productivity? Evidence from a field experiment. Discussion Paper 196, Ludwig-Maximilians-Universität München und Humboldt-Universität zu Berlin.

Ashraf, N., O. Bandiera, E. Davenport, and S. S. Lee (2020). Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *American Economic Review 110*(5), 1355–94.

Ashraf, N., O. Bandiera, and B. K. Jack (2014). No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics 120*, 1–17.

Bai, J., S. Jayachandran, E. J. Malesky, and B. A. Olken (2019). Firm growth and corruption: empirical evidence from vietnam. *The Economic Journal 129*(618), 651–677.

Bandiera, O., I. Barankay, and I. Rasul (2010). Social incentives in the workplace. *The Review of Economic Studies 77*(2), 417–458.

Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (2020). Improving police performance in Rajasthan, India: Experimental evidence on incentives, managerial autonomy and training. *American Economic Journal: Economic Policy*. Forthcoming.

Banerjee, A., E. Duflo, C. Imbert, S. Mathew, and R. Pande (2020). E-governance, accountability, and leakage in public programs: Experimental evidence from a financial management reform in India. *American Economic Journal: Applied Economics 12*(4), 39–72.

Banerjee, A. V. (1997). A theory of misgovernance. *The Quarterly Journal of Economics 112*(4), 1289–1332.

Banerjee, A. V., E. Duflo, and R. Glennerster (2008). Putting a band-aid on a corpse: incentives for nurses in the Indian public health care system. *Journal of the European Economic Association 6*(2-3), 487–500.

Bank, T. W. (2016). *World Development Report 2016: Digital Dividends*. World Bank Publications.

Banuri, S. and P. Keefer (2013). Intrinsic motivation, effort and the call to public service. Policy Research Working Paper 6729, The World Bank.

Barrera-Osorio, F., K. Gonzalez, F. Lagos, and D. J. Deming (2020). Providing performance information in education: An experimental evaluation in colombia. *Journal of Public Economics 186*, 104185.

Becker, G. S. and G. J. Stigler (1974). Law enforcement, malfeasance, and compensation of enforcers. *The Journal of Legal Studies 3*(1), 1–18.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika 93*(3), 491–507.

Bertrand, M., R. Burgess, A. Chawla, and G. Xu (2020). The glittering prizes: Career incentives and bureaucrat performance. *The Review of Economic Studies 87*(2), 626–655.

Bertrand, M., S. Djankov, R. Hanna, and S. Mullainathan (2007). Obtaining a driver's license in india: an experimental approach to studying corruption. *The Quarterly Journal of Economics 122*(4), 1639–1676.

Blader, S., C. Gartenberg, and A. Prat (2020). The contingent effect of management practices. *The Review of Economic Studies 87*(2), 721–749.

Bø, E. E., J. Slemrod, and T. O. Thoresen (2015). Taxes on the internet: Deterrence effects of public disclosure. *American Economic Journal: Economic Policy 7*(1), 36–62.

Bold, T., M. Kimenyi, G. Mwabu, A. Ng'ang'a, and J. Sandefur (2018). Experimental evidence on scaling up education reforms in kenya. *Journal of Public Economics 168*, 1–20.

Callen, M., S. Gulzar, A. Hasanain, M. Y. Khan, and A. Rezaee (2020). Data and policy decisions: Experimental evidence from pakistan. *Journal of Development Economics 146*, 102523.

Cornelissen, T., C. Dustmann, and U. Schönberg (2017). Peer effects in the workplace. *American Economic Review 107*(2), 425–56.

Cowley, E. and S. Smith (2014). Motivation and mission in the public sector: Evidence from the world values survey. *Theory and Decision 76*(2), 241–263.

Dal Bó, E., F. Finan, N. Y. Li, and L. Schechter (2019). Government decentralization under changing state capacity: Experimental evidence from paraguay. Working Paper 24879, National Bureau of Economic Research.

Djankov, S., C. Freund, and C. S. Pham (2010). Trading on time. *The Review of Economics and Statistics 92*(1), 166–173.

Djankov, S., D. Georgieva, and R. Ramalho (2018). Business regulations and poverty. *Economics Letters 165*, 82–87.

Djankov, S., C. McLiesh, and R. M. Ramalho (2006). Regulation and growth. *Economics Letters 92*(3), 395–401.

Dodge, E., Y. Neggers, R. Pande, and C. T. Moore (2018). Having it at hand: How small search frictions impact bureaucratic efficiency. Working paper.

Dustan, A., S. Maldonado, and J. M. Hernandez-Agramonte (2018). Motivating bureaucrats with non-monetary incentives when state capacity is weak: Evidence from large-scale field experiments in peru. Working Paper 136, Peruvian Economic Association.

Finan, F., B. A. Olken, and R. Pande (2017). The personnel economics of the developing state. In *Handbook of Economic Field Experiments*, Volume 2, pp. 467–514. Elsevier.

Freund, C., M. Hallward-Driemeier, and B. Rijkers (2016). Deals and delays: Firm-level evidence on corruption and policy implementation times. *World Bank Economic Review 30*(2), 354–382.

Gerber, A. S., D. P. Green, and C. W. Larimer (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, 33–48.

Hague Institute for Innovation of Law (2018). Justice needs and satisfaction in Bangladesh. Research report, Hague Institute for Innovation of Law.

Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, 74–91.

Huntington, S. P. (1968). *Political Order in Changing Societies*. New Haven: Yale University Press.

Jayaraman, R., D. Ray, and F. De Véricourt (2016). Anatomy of a contract change. *American Economic Review 106*(2), 316–58.

Kaufmann, D. and S.-J. Wei (1999). Does "grease money" speed up the wheels of commerce? Working Paper 7093, National Bureau of Economic Research.

Khan, A. Q., A. I. Khwaja, and B. A. Olken (2016). Tax farming redux: Experimental evidence on performance pay for tax collectors. *The Quarterly Journal of Economics 131*(1), 219–271.

Khan, A. Q., A. I. Khwaja, and B. A. Olken (2019). Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings. *American Economic Review 109*(1), 237–70.

Klapper, L., L. Laeven, and R. Rajan (2006). Entry regulation as a barrier to entrepreneurship. *Journal of Financial Economics 82*(3), 591–629.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects.

The Review of Economic Studies 76(3), 1071–1102.

Leff, N. H. (1964). Economic development through bureaucratic corruption. *American Behavioral Scientist 8*(3), 8–14.

Lewis-Faupel, S., Y. Neggers, B. A. Olken, and R. Pande (2016). Can electronic procurement improve infrastructure provision? Evidence from public works in India and Indonesia. *American Economic Journal: Economic Policy 8*(3), 258–83.

Lui, F. T. (1985). An equilibrium queuing model of bribery. *Journal of Political Economy 93*(4), 760–781.

Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review 99*(1), 112–45.

Maskin, E. and J. Riley (1984). Monopoly with incomplete information. *The RAND Journal of Economics 15*(2), 171–196.

Muralidharan, K. and P. Niehaus (2017). Experimentation at scale. *Journal of Economic Perspectives 31*(4), 103–24.

Muralidharan, K., P. Niehaus, S. Sukhtankar, and J. Weaver (2020). Improving last-mile service delivery using phone-based monitoring. *American Economic Journal: Applied Economics*. Forthcoming.

Mussa, M. and S. Rosen (1978). Monopoly and product quality. *Journal of Economic Theory 18*(2), 301–317.

Myrdal, G. (1968). *Asian drama, an inquiry into the poverty of nations*. London: The Penguin Press.

Nicoletti, G. and S. Scarpetta (2003). Regulation, productivity and growth: OECD evidence. *Economic Policy 18*(36), 9–72.

Niehaus, P. and S. Sukhtankar (2013a). Corruption dynamics: The golden goose effect. *American Economic Journal: Economic Policy 5*(4), 230–69.

Niehaus, P. and S. Sukhtankar (2013b). The marginal rate of corruption in public programs: Evidence from India. *Journal of Public Economics 104*, 52–64.

Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in indonesia. *Journal of Political Economy 115*(2), 200–249.

Perez-Truglia, R. and U. Troiano (2018). Shaming tax delinquents. *Journal of Public Economics 167*, 120–137.

Rasul, I. and D. Rogger (2018). Management of bureaucrats and public service delivery: Evidence from the nigerian civil service. *The Economic Journal 128*(608), 413–446.

Reinikka, R. and J. Svensson (2005). Fighting corruption to improve schooling: Evidence from a newspaper campaign in uganda. *Journal of the European Economic Association 3*(2-3), 259–267.

Rose-Ackerman, S. (1978). *Corruption: A study in political economy*. New York: Academic Press.

Rosenzweig, M. R. and C. Udry (2020). External validity in a stochastic world: Evidence from low-income countries. *The Review of Economic Studies 87*(1), 343–381.

Singh, A. (2020). Myths of official measurement: Auditing and improving administrative data in developing countries. Working Paper 20/042, RISE.

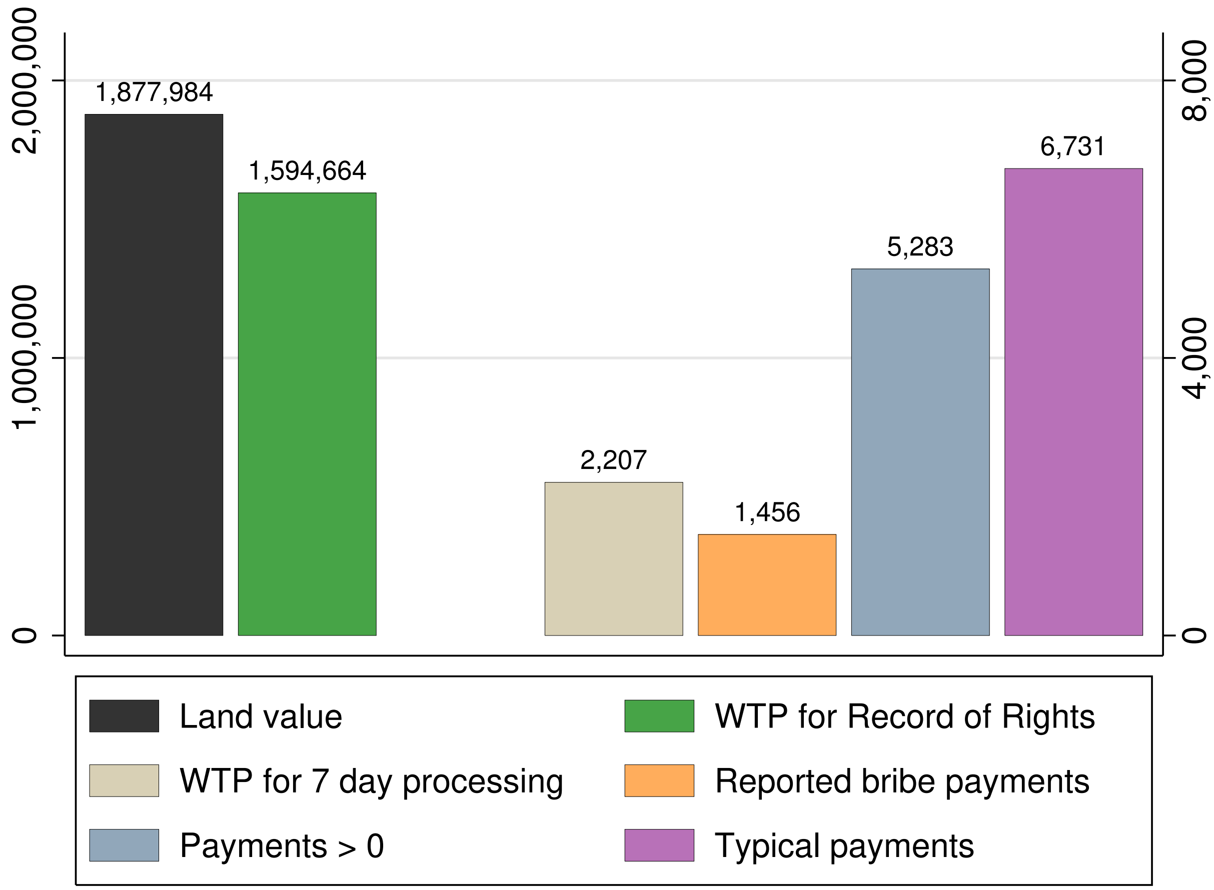
Svensson, J. (2003). Who must pay bribes and how much? Evidence from a cross section of firms. *The Quarterly Journal of Economics 118*(1), 207–230.

Transparency International Bangladesh (2016, 06). Corruption in service sectors, national household survey 2015. Research report, Transparency International Bangladesh.

Transparency International Bangladesh (2018). Corruption in service sectors: National household survey 2017. Research report, Transparency International Bangladesh.

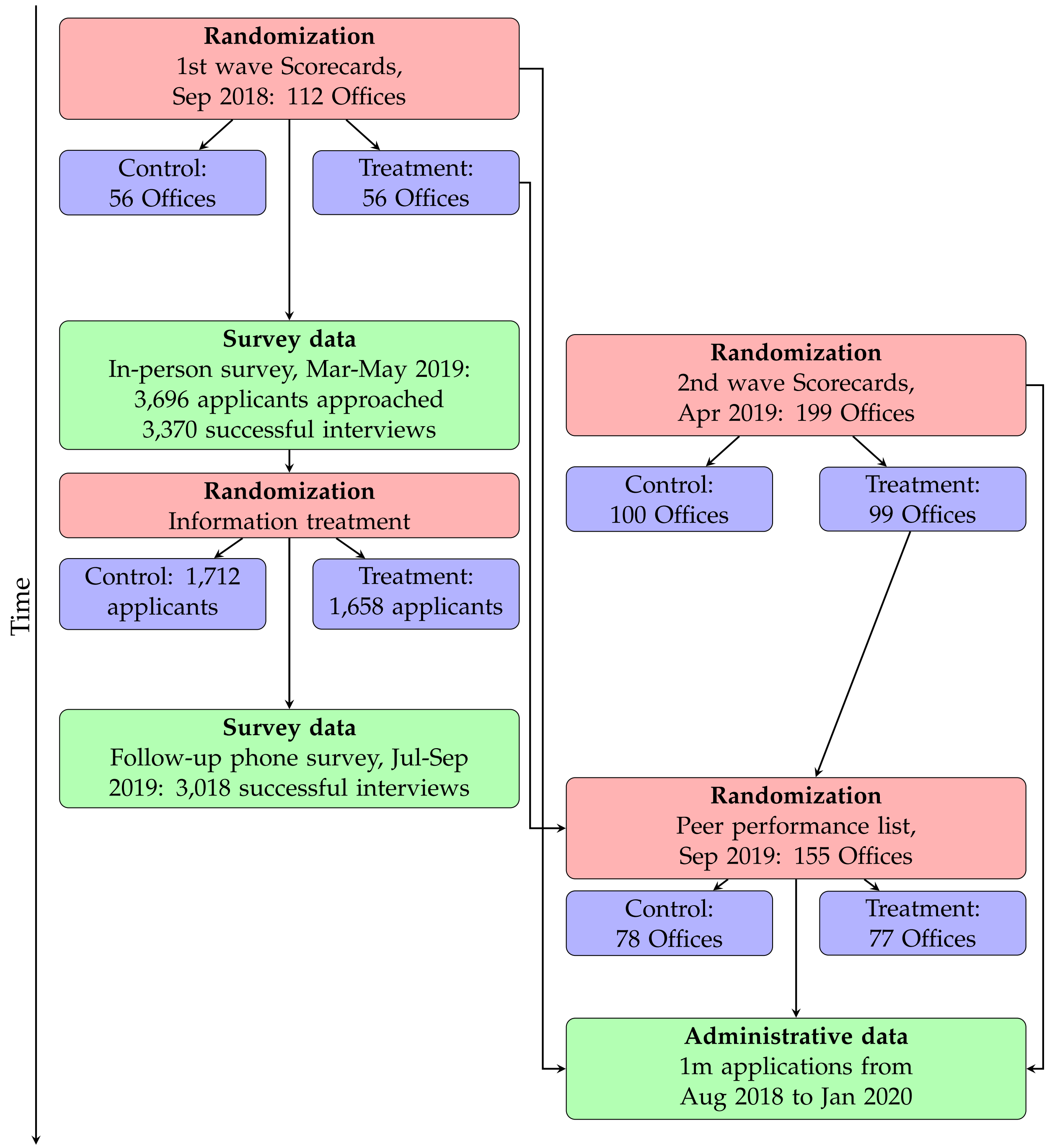
Weaver, J. (2020). Jobs for sale: Corruption and misallocation in hiring. Working paper.

Figure 1: Value of land, record of rights, faster processing, and bribe payments



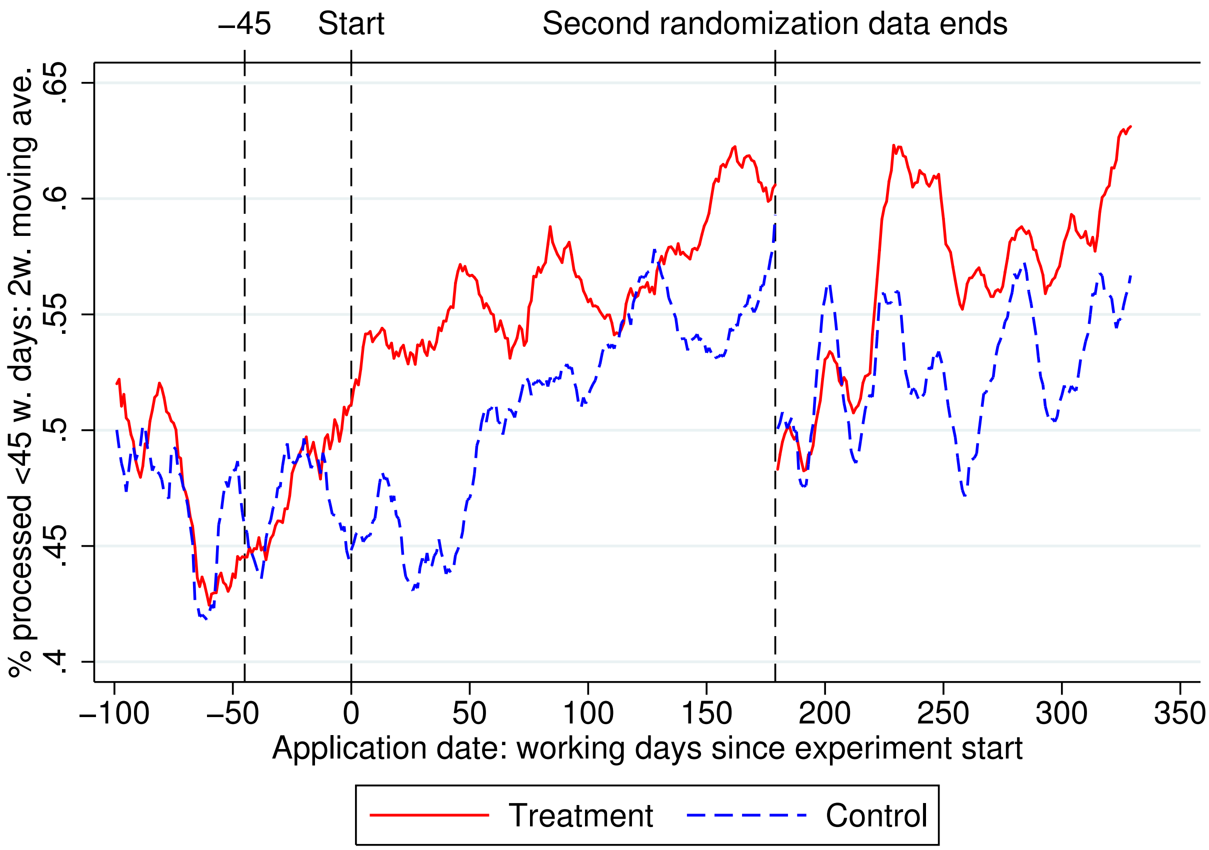
This figure shows the average value of land, the applicants' stated valuation of the record of rights, the applicants' valuation of getting their application processed within seven days, and three different ways of measuring average bribe payments. All variables are winsorized at the 99th percentile. Observations are weighted by the inverse of the number of observations in their land office. The first bar shows the average of applicants' estimates of the value of the land for which the land record change is being applied for. The second bar shows the average of applicants' stated value of getting the land record change approved and receiving a record of rights. The third bar shows the average stated value of getting the application processed within seven days from the time of the first survey. The fourth bar shows the average value of bribe payments reported by the applicant, 73% of the applicants reported having paid no bribes. The fifth bar shows the average value of bribe payments reported by applicants reporting having paid some bribe. The sixth bar shows the average value of an estimated "typical bribe payment by a person like yourself" reported by the applicant, 27% of the applicants responding to this question reported that a typical applicant paid no bribes. The first two bars are measured on the axis on the left, the next four bars are measured on the axis to the right. Discussed in Sections 2.1 and 6.

Figure 2: Overview of randomization and data collection



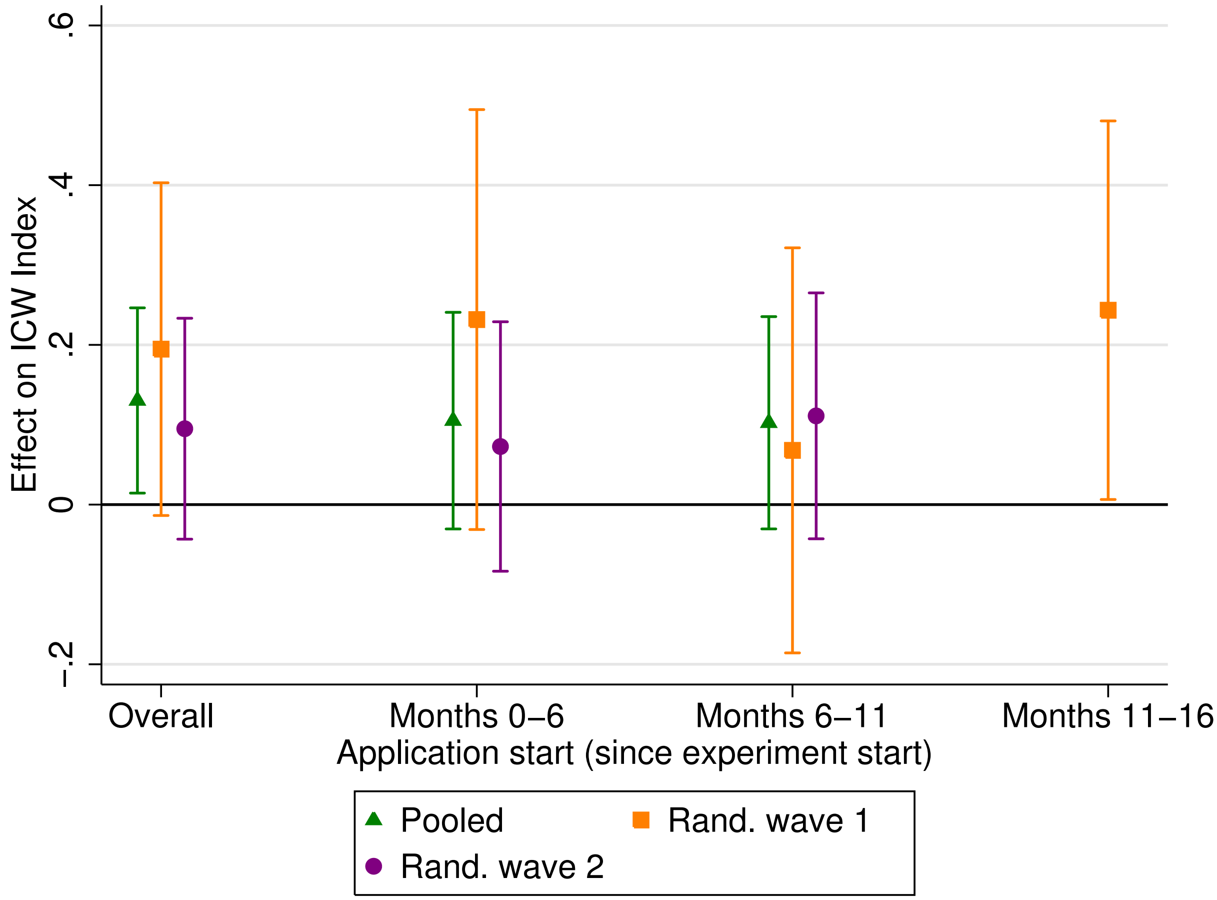
This figure provides a visual overview of the experiment design and data collection. Boxes places further down in the figure represent things that happened later with the exception of the administrative data collection, which happened throughout the project. Red boxes represents randomizations into treatment or control. Blue boxes represents the treatment and control groups. Green boxes represents data collection. Discussed in Section 2.4.

Figure 3: Fraction of applications processed within 45 working days



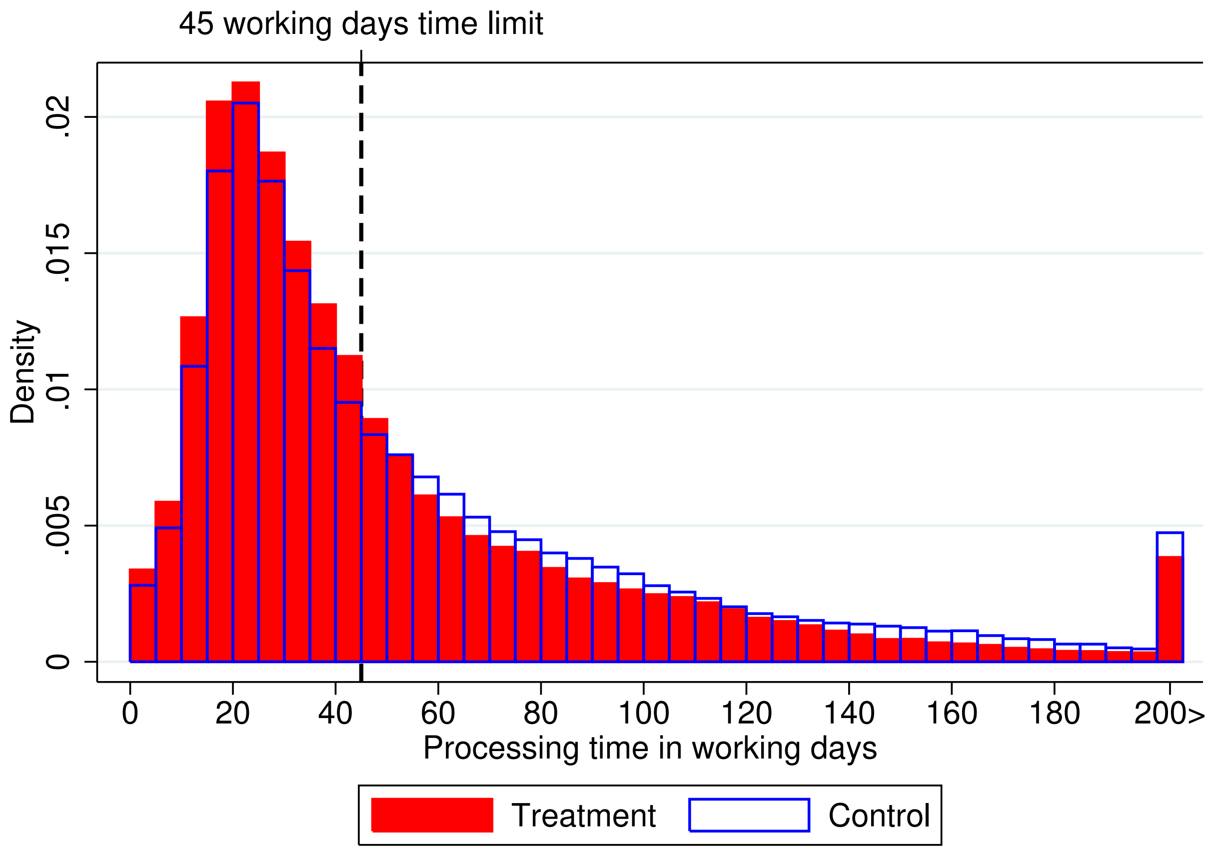
This figure shows the two week moving average of the daily fraction of applications processed within the 45 working day limit in the treatment and control groups. Data contains all applications made from 100 working days before the start of the experiment until 45 working days before the experiment ended (25 Apr 2018 - 20 Jan 2020). The first vertical line represents the date 45 working days before the first scorecard was sent out. In principle, the effect can have started for any applications made after this date. The second vertical line represents the date of the first scorecard, applications made after this date were fully treated. The third vertical line represents the end of the data from the second randomization wave. To the right of this line the figure is based on the 112 offices in the first randomization wave. Discussed in Section 4.1.1.

Figure 4: Estimated effect of scorecards on processing times by time since the start of experiment



This figure shows the regression coefficients and confidence intervals for regressions using applications started during different time periods after the experiment started. The outcome variable is the ICW Index from Column (3) of Table 2. The overall estimate uses the same specification as in Table 2. The estimates for the three time periods are estimated by interacting a dummy variable for if the application was made in that time period with the treatment variable. The figure show results from regressions using data from all offices (triangle), offices in the first randomization wave (squares), and offices in the second randomization wave (circles). The months are numbered relative to when scorecards were sent out for that office's randomization wave. Month 0 is the month before the start of the experiment. Confidence intervals are constructed using standard errors clustered at the office level. Discussed in Section 4.1.1.

Figure 5: Histogram of processing times by treatment



This figure shows histograms of processing times for the treatment and control groups separately. Processing times are top coded at 200 working days. Data contains all applications made between 1 month before the start of the experiment and 45 working days before the experiment ended (13 Aug 2018 - 20 Jan 2020). Applications not yet processed are excluded. Discussed in Section 4.1.2.

Table 1: Summary statistics

	(1) Mean	(2) Median	(3) St. Dev.	(4) Observations
Panel A: Application level administrative data				
Process time < 45 w. days	0.59	1	0.49	1,050,924
Actual process times (w. days)	50	34	45	972,589
Process time inc. imputed values (w. days)	63	36	70	1,050,924
Approval rate	0.69	1	0.46	972,582
Panel B: Monthly office level administrative data				
Total applications	287	213	272	4,516
Applications processed	240	136	331	4,516
Apps. disposed within 45 w. days	150	79	195	4,516
Apps. pending beyond 45 w. days	382	97	732	4,516
Panel C: Applicant survey data				
Applicant age	47	47	14	2,903
Female	0.06	0	0.24	3,018
Applicant monthly income (BDT)	23,902	20,000	21,093	2,791
Applicant HH per capita expenditure (BDT)	4,400	3,462	3,537	3,018
Land Value (BDT 100,000)	19	8	30	2,800
Land Size (Decimal = 1/100th Acre)	24	10	40	2,892
Any additional payment made	0.28	0	0.45	3,018
Reported payment amount (BDT)	1,456	0	3,456	3,018
Typical payment amount (BDT)	6,731	5,000	8,414	1,896

This table shows summary statistics for applications in the administrative data, offices, and applicants in the survey data. Observations in Panel A and C are inversely weighted by the number of applications in their land office. Observations in Panel B are uniformly weighted. Continuous variables in the survey data are winsorized at the 99th percentile. USD/BDT≈84.3. Reported payment amount is any payment reported by the applicant above the official fee. Typical payment amount is the answer to the question of my much a "normal person, like yourself" typically pay to get an application processed. Discussed in Section 2.5.

Table 2: Effect of scorecards on processing times

	(1) <45 w. days	(2) IHS(w. days)	(3) ICW index
Scorecard	0.0608**	-0.125**	0.130**
	(0.0275)	(0.0593)	(0.0592)
Start month FE	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes
Observations	1,050,924	1,050,924	1,050,924
Clusters	311	311	311
Control mean	0.56	65.64	-0.00
Fraction imputed		0.06	
Fraction zero		0.003	

This table shows the effect of the scorecards on the speed of application processing. Column (1) shows the effect on the fraction of applications processed within the 45 working day time limit. Column (2) shows the effect on the IHS transformation of processing time. Applications that are not yet processed are given imputed processing times equal to the mean of processing times that are longer than the application has currently been pending. Column (3) shows the effect on an inverse covariance weighted matrix combining the outcome variables of Columns (1) and (2). Data contains all applications made between 1 month before the start of the experiment and 45 working days before the experiment ended (13 Aug 2018 - 20 Jan 2020). Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in their land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 4.1.

Table 3: Effect of peer performance list

	(1) <45 w. days	(2) IHS(w. days)	(3) <45 w. days	(4) IHS(w. days)
Peer Performance List	0.00483	-0.0360	0.0116	-0.0404
	(0.0452)	(0.0948)	(0.0394)	(0.0823)
Scorecard			0.0570*	-0.112*
			(0.0293)	(0.0651)
Start month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	286,152	286,152	1,050,924	1,050,924
Clusters	155	155	311	311
Control mean	0.67	51.31	0.59	65.64
Fraction imputed		0.06		0.06
Fraction zero		0.003		0.003

This table shows the effect of adding the list of performances to the scorecard and sharing the performance information about one office with the ACLs, UNOs, and DCs of 76 other offices. Column (1) shows the effect on the fraction of applications processed within the 45 working day time limit. Column (2) shows the effect on the IHS transformation of the processing time. Columns (3) and (4) show the effects of both the scorecard treatment without the performance list and the performance list separately. Columns (1) and (2) only use data from offices in the performance list experiment and data from applications made one month before the first performance list until 20 January 2020. Columns (3) and (4) use data on all applications made between 1 month before the start of the experiment started and 45 working days before the experiment ended (13 Aug 2018 - 20 Jan 2020). The dummy variable "Peer performance list" takes the value of one for applications made in offices receiving the peer performance lists, made later than one calendar month before the first performance list was sent out. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 4.2.

Table 4: Effect on visits and time spent by applicants

	(1) Visits	(2) Hours spent	(3) ICW index
Scorecard	-1.034**	-1.586	0.0851
	(0.497)	(1.855)	(0.0552)
Start month FE	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes
Observations	3,018	3,018	3,018
Clusters	112	112	112
Control mean	8.99	23.66	

This table shows the effect of the scorecards on visits to land offices and the number of hours spent on these visits. Standard errors are clustered at the land office level. Observations inversely weighted by the number of applications in the land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 4.3.

Table 5: Effect on bribe payments for application processing

	Amount		Any bribe		Amount if > 0	
	(1)	(2)	(3)	(4)	(5)	(6)
Scorecard	1,046*	265	-0.014	-0.003	1,573**	1,069**
	(615)	(181)	(0.022)	(0.022)	(768)	(457)
Start month FE	Yes	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,896	3,018	1,896	3,018	1,392	807
Clusters	112	112	112	112	112	111
Control mean	6,083	1,278	0.75	0.27	8,083	4,700
Bribe measure	Typical	Reported	Typical	Reported	Typical	Reported

This table shows the effect of the scorecards on bribe payments made for application processing. Column (1) shows the effect on the estimate for how much a "normal person, like yourself" pays in bribes to process an application. Column (2) shows the effect on reported payments to government officials or agents beyond the official fee. Columns (3) and (4) show the effect on the fraction of non-zero answers for the two questions. Columns (5) and (6) show the effect among applicants who reported a non-zero bribe. All monetary amounts are in BDT. USD/BDT≈84.3. All continuous variables are winsorized at the 99th percentile. Standard errors are clustered at the office level. Observations inversely weighted by the number of applications in the land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 4.4.

Table 6: Effect on processing times, visits, and time spent for offices over-performing and under-performing at baseline

	(1) <45 w. days	(2) IHS(w. days)	(3) Office visits	(4) Hours spent
Scorecard x Overperform baseline	0.00823	-0.0345	-0.678	0.358
	(0.0372)	(0.0804)	(0.729)	(2.626)
Scorecard x Underperform baseline	0.124***	-0.234***	-1.230*	-2.973
	(0.0402)	(0.0876)	(0.731)	(2.843)
Overperform baseline	0.193***	-0.315***	-1.854*	-7.095**
	(0.0502)	(0.108)	(0.949)	(3.318)
P-value sub-group diff.	0.04	0.10	0.10	0.30
Start month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	1,050,924	1,050,924	3,018	3,018
Clusters	311	311	112	112
Overperformers: q-value	1.00	1.00	0.55	1.00
Underperformers: q-value	0.02	0.03	0.24	0.55
Overperformers: control mean	0.68	51.53	8.03	20.37
Underperformers: control mean	0.41	82.54	9.88	26.72

This table shows the effect of the scorecards separately for offices with above- and below-median performance at baseline. In Columns (1) and (2), results are based on administrative data. In Columns (3) and (4), results are based on survey data. Column (1) shows the effects on the fraction of applications processed within the 45 working day limit. Column (2) shows the effects on the IHS transformation of the number of working days it took to process the application. Column (3) shows the effect on the number of visits to land offices needed for the processing of the application. Column (4) shows the effect on the number of hours spent by the applicant for the processing of the application. Standard errors are clustered at the land office level. Q-values are sharpened false discovery rate q-values for the eight hypotheses that the effect of the scorecards is zero for all outcome variables and for both over-performers and under-performers (Benjamini, Krieger, and Yekutieli, 2006; Anderson, 2008). Observations inversely weighted by the number of applications in the land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 4.5.1.

Table 7: Effect on bribes for offices over-performing and under-performing at baseline

	(1) Typical payment	(2) Reported payment
Scorecard x Overperform	2279.7***	638.4***
	(772.9)	(229.1)
Scorecard x Underperform	-162.4	-59.38
	(954.6)	(257.8)
Overperform baseline	-1811.8*	-819.0***
	(928.4)	(287.8)
P-value sub-group diff.	0.06	0.05
Start month FE	Yes	Yes
Stratum FE	Yes	Yes
Weighted by office	Yes	Yes
Observations	1,896	3,018
Clusters	112	112
Overperformers: q-value	0.01	0.01
Underperformers: q-value	0.76	0.76
Overperformers: control mean	5,313	916
Underperformers: control mean	6,817	1,616

This table shows the effect of the scorecards on offices with above- or below-median performance at baseline. Column (1) shows the effects on what the applicant reports to be a typical payment for a land record change for a person like themselves. Column (2) shows the effects on the payments reported by the applicant. The outcome variables are in BDT. USD/BDT≈84.3. Standard errors are clustered at the land office level. Q-values are sharpened false discovery rate q-values for the four hypotheses that the effect of the scorecards is zero on both outcome variables and for both over-performers and under-performers (Benjamini et al., 2006; Anderson, 2008). Observations are inversely weighted by the number of applications in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 4.5.2.

Appendix

A Theory

A.1 Model of reputation and bureaucrat behavior

A.1.1 Model set-up

Government bureaucrats utility function:

$$U\left(E_i, B_i, t_i R(B_i, v^{T_i} P(E_i))\right) = D(E_i) + M(B_i) + t_i R\left(B_i, v^{T_i} P(E_i)\right) \tag{3}$$

- Subscript i represent individual bureaucrats
- $D(E)$ is disutility of effort E, which has negative first and second derivatives, $D'(E) < 0$ and $D''(E) \leq 0$
- $M(B)$ is the utility of bribe money B, which has a positive first derivative and negative second derivative, $M'(B) > 0$ and $M''(B) \leq 0$
- $R(B, vP(E))$ is the utility from reputational concerns which is determined by B and visible performance $vP(E)$
  - $R(.)$ has a negative derivative with respect to bribes $R^1(B, vP(E)) < 0$ and a positive derivative with respect to visible performance $R^2(B, vP(E)) > 0$. Both second derivatives are negative, $R^{11}(B, vP(E)) < 0$ and $R^{22}(B, vP(E)) < 0$
  - The cross derivative is positive, $R^{21}(B, vP(E)) > 0$, i.e., bribes and performance are complements, or equivalently honesty (the lack of bribes) and performance are substitutes
  - A technical assumption used to ensure the existence of derivatives and avoid corner solutions is $R^{12}(B, vP(E)) \leq \left(R^{11}(B, vP(E)) R^{22}(B, vP(E))\right)^{\frac{1}{2}}$
- v^{T_i} is the visibility of performance $P(E_i)$ and depend on bureaucrat i's treatment $T_i \in \{scorecard, control\}$ such that $v^{scorecard} > v^{control}$
  - Connecting the performance term directly to the scorecards, I assume that performance is the average ranking of a bureaucrat in terms of applications processed on time and applications pending longer than the time limit

- $P(E)$ is increasing in E, positive when E is above median effort and negative when E is below median effort
  - $*$ The second derivative is zero or negative, $P''(E) \leq 0$

- t is the type of bureaucrat and reflects the degree to which the bureaucrat values reputation
  - Bureaucrats only differ in their valuation of reputation t and their treatment status v

All of the assumptions, including the technical assumptions mentioned below, are fulfilled by a simple Cobb-Douglas utility function of the form:

$$U = \alpha ln(1-E) + \beta ln(B) + tln(c + v(E - \bar{E}) - B)$$

Where c is a constant sufficiently large so that $c + v(E - \bar{E}) - B > 0$ and $\bar{E}$ is the effort of the median bureaucrat when $v^T = v^{control}$.

For simplicity, I do not formally model applicants' behavior but assume that they have no choice but to accept the bureaucrats' bribe request.

A.1.2 Solution to bureaucrats problem

Bureaucrats choose E and B to maximize $U(E, B, tC(B, vP(E)))$. The first order conditions to the bureaucrats maximization problem are:

$$D'(E_i^*) + t_i R^2\left(B_i^*, v^T P(E_i^*)\right) v^T P'(E_i^*) = 0 \tag{4}$$

Where E^* and B^*represent the choices of E and B that maximize utility for bureaucrat i. At the optimum, the marginal disutility of effort is equals the marginal utility of effort's effect on reputational concerns.

$$M'(B_i^*) + t_i R^1\left(B_i^*, v^T P(E_i^*)\right) = 0 \tag{5}$$

At the optimum, the marginal utility of bribe money equals the marginal disutility from a decrease in reputation due to bribes.

A.1.3 Effect of scorecards on effort

Henceforth, I will drop the star superscript (*) on E and B since all mentions will refer to the values of E and B at the optimum. Taking the total derivatives of the first order conditions with respect to v gives us the following expression for the derivative of E with respect to v:

$$\frac{dE_i}{dv} = P'(E_i)\frac{-\left(\frac{M''(B_i)}{t_i} + R^{11}(.)\right)R^2(.) + \left(\left(R^{12}(.)\right)^2 - \left(\frac{M''(B_i)}{t_i} + R^{11}(.)\right)R^{22}(.)\right)v^T P(E_i)}{\left(\frac{D''(E_i)}{t_i} + R^{22}(.)\left(v^T P'(E_i)\right)^2 + R^2(.)\,v^T P''(E_i)\right)\left(\frac{M''(B_i)}{t_i} + R^{11}(.)\right) - \left(R^{12}(.)\,v^T P'(E_i)\right)^2} \tag{6}$$

Using the technical assumption $R^{21}(B, vP(E)) \leq \left(R^{11}(B, vP(E))\,R^{22}(B, vP(E))\right)^{\frac{1}{2}}$ the denominator is positive. The first term in the numerator reflects the direct substitution effect of the changed visibility of effort. This effect is always positive. The second term in the numerator represents the income effect, i.e., the effect on effort stemming from improved reputation because of the change in visibility of existing efforts. If $P(E) < 0$ then this term is also positive making the whole expression positive.

Prediction 1: If $P(E) < 0$ then $\frac{dE}{dv} > 0$. I.e., the scorecards increase effect and therefore improve processing times for offices under-performing at baseline

For bureaucrats with an above-median performance, when their positive performance becomes more visible, their reputation improves, and the marginal utility from exerting effort on improving reputation decreases. For bureaucrats with an above-median effort, the effect is, therefore, ambiguous.[34]

A.1.4 Effect of scorecards on bribes

Taking the total derivatives of the first order conditions with respect to v gives us the following expression for the derivative of B with respect to v:

$$\frac{dB_i}{dv} = R^{12}(.)\frac{v\left(P'(E_i)\right)^2 R^2(.) - \left(\frac{D''(E_i)}{t_i} + 2R^{22}(.)\left(v^T P'(E_i)\right)^2 + R^2(.)\,vP''(E_i)\right)P(E_i)}{\left(\frac{M''(B_i)}{t_i} + R^{11}(.)\right)\left(\frac{D''(E_i)}{t} + R^{22}(.)\left(vP'(E_i)\right)^2 + R^2(.)\,vP''(E_i)\right) - \left(R^{21}(.)\,vP'(E_i)\right)^2} \tag{7}$$

Again using the technical assumption $R^{21}(B, vP(E)) \leq \left(R^{11}(B, vP(E))\,R^{22}(B, vP(E))\right)^{\frac{1}{2}}$ the denominator is positive. The first term in the numerator is positive and derived from the substitution effect

[34] The ambiguous effect is analogous to the effect on labor supply from a wage increase. The income effect and price effect go in different directions and depending on which dominates the overall effect may be positive or negative.

increasing effort, which leads to an improvement in visible performance which in turn leads to an increase in bribes because of the complementarity between visible performance and bribes. The second term is derived from the changed visibility of pre-existing performance which also affects bribes due to the complementarity between visible performance and bribes. For bureaucrats with an effort above the median effort this effect is positive. Their positive performance becomes more visible so their reputation term improve and the complementarity decreases the marginal disutility, through the reputation channel, from collecting bribes. Conversely, for bureaucrats with below-median effort this effect is negative. Hence, for bureaucrats with above-median effort the scorecards leads to higher bribes ($\frac{dB}{dv} > 0$) while for bureaucrats with below-median effort the effect is ambiguous.

Prediction 2: If $P(E) > 0$ then $\frac{dB}{dv} > 0$. I.e., scorecards increase bribes for offices over-performing at baseline

A.2 Monopolistic price discrimination model connecting service delivery speed and bribes

The experiment was designed to test a specific model of how the speed of application processing and bribes are connected. A full exposition of the complete model and its predictions is available in the pre-analysis plan. Here I outline the intuition behind the model and its prediction for the experiment. The model is based on an asymmetric information model of price discrimination under monopoly where the bureaucrat acts as monopolists selling a service. Applicants get utility from having their application processed, the faster the application is processed the more utility the processing generates. Applicants only differ only in their willingness to pay for the speed of processing their applications. All applicants' utility is linear in money and non of them are liquidity constrained so their willingness to pay equals their ability to pay. The bureaucrat can ask for different bribe payments from the applicants and can offer the service with different processing times or refuse to provide the service. Once a processing time and bribe payment is agreed upon the applicant pays the bribe and the bureaucrat must honor the agreement. The bureaucrat gets utility from receiving bribes. It can be costly for the bureaucrats to process the application faster although this is not necessary for the main conclusions of the model.

Perfect information in the context of this model means that the bureaucrat can perfectly observe the applicants' willingness to pay for having the application processed faster. Under perfect information, applicants will have their applications processed at a Pareto optimal speed where the marginal benefit of having the application processed faster is the same as the marginal cost of processing the application faster for the bureaucrat. Asymmetric information means that bureaucrats cannot observe the applicants' willingness to pay. Under asymmetric information, the bureaucrat has to offer the same menu of processing times and bribe payments to all applicants.[35] Under asymmetric information only the applicants with the highest willingness get their application processed at the Pareto optimal speed. All other applicants have their applications slowed down as the bureaucrats trade-off providing fast processing for applicants with lower willingness to pay with how large of a bribe they can charge from applicants with a higher willingness to pay.

An easy way to see this trade-off is in a simple example where it is costless for the bureaucrat to process the application immediately and there are two types of applicants, one with a higher willingness to pay to have the application processed quickly. Under full information, the bureaucrat can simply make a take-it-or-leave-it offer to all applicants at exactly their willingness to pay to have the application processed immediately. The applicants will pay their respective willingness to pay because they have no better outside option and the bureaucrat will process the applications immediately. Under asymmetric information, the bureaucrat cannot differentiate between the applicants ex-ante. It now becomes optimal, from the bureaucrat's perspective, to offer to process the application immediately at a higher bribe payment and slower at a lower bribe payment. The applications for those with low willingness to pay are now intentionally delayed despite that processing them immediately does not cost the bureaucrat anything.

A.2.1 Predictions of the model for experiment

The scorecards encourage bureaucrats to process applications within 45 working days. Section 4 shows that the scorecards led to an increase in the applications processed within 45 working days and that the effect was driven by offices that were under-performing at the start of the experiment.

Under full information, an increase in processing speed is predicted to lead to a slight increase in bribe payments among those whose applications are processed faster. This is because for these applications the value of the processing has increased and they are now willing to pay more for it. Under asymmetric information, an increase in processing speeds is predicted to reduce the bribe payments among those with the highest willingness to pay for getting their applications processed quickly. This is because the bureaucrat has to make the menu option of having applications being processed quickly more attractive in order for these applicants to continue to pay for it now that the processing speed of

[35] Realistically, some observable characteristics contain information about the applicants' willingness to pay. In this case, the bureaucrat has to offer the same menu to all applicants with the same observable characteristics.

the option to pay less has increased.

A.2.2 Testing predictions from monopolistic price discrimination model

The experiment was set up to test a specific maximizing bribes model, described in Appendix Section A.2. The main testable prediction of this model with respect to the experiment is that when delays in processing times are reduced, the bribe payments by those who have a high willingness to pay for getting their application processed quickly should decrease. This result stems from that the bureaucrat uses the difference in terms of processing times between those paying large and small bribes, to maintain a separating equilibrium and maximize the amount of bribes extracted. In particular, if long delays were reduced, but the bribes for those who paid the largest bribes and subsequently got the fastest processing time did not change, then some of these applicants would choose to pay a lower bribe and get a slower processing time. Anticipating this the bureaucrats would reduce the bribes for those with the highest willingness to pay for fast services and thereby maintain the separating equilibrium while still providing efficiently fast services for the applicants with the highest willingness to pay. The results from the experiment are inconsistent with this prediction.

Column (1) of Appendix Table A3 shows the effect on bribes among those who had their applications processed quickly, using 25 working days as the cutoff for if an application was processed quickly. Column (1) shows the effect was positive and that the scorecards increased the bribe payments among applicants who had their applications processed within 25 working days by BDT 656. Column (2) shows that for applications processed outside of 25 working days limit the estimated effect was a BDT 333 increase but this effect is not statistically significant. Column (3) shows that even for offices that were under-performing at baseline, that had the largest decrease in delays and processing time as a result of the scorecards, the effect of the scorecards on bribes for applications processed within 25 working days is estimated to be an increase of BDT 686. Although the effect is not statistically significantly different from zero, any meaningful negative effect can be rejected.

One potential explanation for the results within the framework of the monopolistic price discrimination model is that the government officials taking the bribes have full information about the applicants' willingness to pay for processing speed. In this information setting the speeding up of processing for those with lower willingness to pay would not affect those with higher willingness to pay, since the bureaucrat could maintain the separating equilibrium by simply requesting different bribes depending on the observed willingness to pay. However, even under full information, the bribe payments are not predicted to increase for those with the highest willingness to pay. The effects of the scorecards on bribes shown in Appendix Table A3 are therefore not consistent with the predictions of the model under any information setting.

Another explanation for why the results are inconsistent with the predictions of the model is that information about the increase in processing speeds had not yet been disseminated to applicants by the time of the survey period. The information treatment is designed to alleviate this problem. Column (4) of Appendix Table A3 shows that the information treatment had no effect on bribes by itself or in combination with the scorecard treatment. Finally, Column (5) shows that even for applicants that received the information in offices that were under-performing at baseline, no negative effect on bribes can be found.

Taken together the results from the experiment rejects the model's predictions.

B Additional Details on Experiment and Data

B.1 Details on randomization of scorecards treatment assignment

The randomization of which offices were assigned to receive the scorecards was done at the land office level. The first wave randomization was done separately for the group of land offices classified by the Government as having full implementation of the e-governance system at the start of the experiment and for the group with partial implementation at the start of the experiment. After these two groups had been separated the the randomization strata were created using the following variables:

- Number of applications processed within 45 working days in the months of June and July 2018
- Number of applications pending for more than 45 working days at the end of July 2018

For the group of offices with partial e-governance implementation, stratum were created based on offices being in the first, second or third tertile in the distribution of these variables. Since the group of offices with full e-governance implementation was smaller, stratum were created based on offices being above or below the median in the distribution of these variables.

The second wave randomization was done separately for the group of land offices having received above/below the median number of applications in February and March 2019. After these two groups had been separated the randomization strata were created using the following variables:

- Being in the first, second or third tertile in terms of number of applications processed within 45 working days in March 2019

- Being in the first, second or third tertile in terms of applications pending for more than 45 working days at the end of March 2019

Within each strata, half of the land offices were randomly assigned to treatment.[36] If there was an odd number of offices in a strata, the last office is grouped together with other such "misfits" in their implementation group (first wave) or applications received group (second wave) and half of the misfits were randomly assigned treatment. Again, when there were misfits these are grouped together with other misfits from the other implementation group and half of those were assigned treatment. Finally, the last misfits were assigned treatment with a 50% probability.

B.2 Data

B.2.1 Administrative data from e-governance system

The government partner transferred the data at the beginning of each month from August 2018 until September 2020. Due to privacy concerns the Government only shared administrative data without personal identifying information. The administrative data is at the application level and includes all applications made in the e-governance system since its inception in February 2017. To calculate the number of working days between the date the application started until the end of the application I use data on public, national and general holidays from Time and Date. [37]

Although the performance scorecards are addressed to the current ACL of a land office, the performance is based on how applications made in that land office are processed, regardless of if that ACL was assigned to that office when the application was made or not. I deduct what ACL was assigned to what office using the administrative data on what ACL made updates on the applications in a given month. I use administrative user data to separate ACLs from other users and then assign a particular ACL to an office if that ACL is the ACL making the largest number of updates in an office in a particular month. If an office has no updates made by any ACL in a month, I do not assign any ACL to that office in that month, unless there is an ACL that was assigned to that office both prior to and after that month, in which case I deduct that the ACL was assigned to that office without making any updates in that month.

It would be very difficult for any individual bureaucrat to improve their performance scorecard by manipulating the administrative data. The data is stored on a central server that they bureaucrats do not have access to. While it would be possible to create fake applications in the e-governance system, to

[36] Random assignment was implemented by the author using the Stata command runiform.
[37] As of September 2020, available from here: https://www.timeanddate.com/holidays/bangladesh/

process these applications with an acceptance the processing fee would have to be paid. Hence, creating fake applications would decrease, not increase, a bureaucrat's performance ranking.

During training, several example applications were made in two land offices were the e-governance system was not yet installed. These "fake" applications were then never removed from the system making it appear as if the e-governance system was active in these two offices. Due to this these offices were included in the first wave of randomization, one was assigned treatment and one control. In September 2018 I found out that these two offices had not yet installed the e-governance system and I removed all applications from these offices from the administrative data and stopped sending the scorecards to the office that had been assigned the treatment. Some other applications in the e-governance system are also the result of examples created in training. Using information provided about the dates of the training, I removed applications made before the first wave of randomization suspected to be the result of training. I did not remove any applications made after the start of the experiment.

B.2.2 Survey data

The survey was carried out in two stages. The average time between the two interviews was 3.3 months. All questions about bribe payments were asked in the second, interview made by phone. Interviewees were given a BDT 50 (USD 0.6) reward in the form of a mobile phone recharge for a completed in-person interview and BDT 100 (USD 1.2) for a completed phone interview.

B.2.3 Attrition in the survey data

The attrition rate was 18%. Appendix Table A13 estimate the effect of the scorecard and information treatments on the attrition rate. Column (1) shows that the scorecard treatment is estimated to have had a positive effect on the attrition rate by 3%, an effect which is statistically significant at the 10% level. Columns (2) and (3) show that the information intervention did not affect attrition. Column (4) shows that the effect of the scorecards on attrition is mainly concentrated among offices under-performing at baseline.

If the scorecards caused some applicants to drop out of the study and these applicants, on average, had different values for an outcome variable, this would bias the estimates of the effect of the scorecards on those outcomes. To assess the potential bias stemming from the differential attrition on the estimated effect of the scorecards on bribe payments, I construct lower Lee bounds for the estimated effect (Lee, 2009). Lower Lee bounds are the relevant robustness check, since the effects on bribe payments are positive (overall and for over-performing offices) or non-negative (for under-performing offices). I create lower Lee bounds by creating a random selection of the applicants from treated offices for whom there is no follow-up survey data, the random selection is equal in size to the estimated effect of the scorecards on the number of applicants not completing the follow-up interview. I then set the bribe payments for this sub-sample to zero, since that is the lowest possible bribe payment. I then conduct the main analysis from Column (1) of Table 5 and Column (1) of Table 7. The results are shown in Appendix Table A14. The lower Lee bounds does not qualitatively change the overall results. The estimated effect of the scorecards on bribes is BDT 205. For offices over-performing at baseline the estimated lower Lee bound effect is BDT 639 and still statistically significant. For offices under-performing at baseline the estimated lower Lee bound is BDT -177. The Lee bounds show that even if the entire effect of the scorecards on attrition was on applicants who would have reported zero bribes, this would not have substantially changed the results.

B.2.4 Cross-validation of administrative and survey data on processing times

There are two application numbers identifying the applications making it possible to match applications from the survey with applications in the administrative data. One is a global identification number automatically generated by the e-governance system. Another is a manually generated local serial number, unique within a year and a land office. The local serial number can also be entered into the e-governance system, but is not a required field and is hence missing for some applications. Unfortunately, few applicants shared their digital application ID, either because they were uncomfortable giving it out or because they could not find the text message through which they received this number. There were many inconsistencies in how the local serial numbers were recorded between and within offices. There also exist several other serial numbers, such as a serial number for the record of rights, that could be confused with the application serial number and that were sometimes reported in the survey instead of the application serial number. This caused many duplicate serial numbers, even within an office by year combination. Due to these problems, only 45% of the applications from the survey that could be matched to the administrative data. Furthermore, it is possible that incorrect matches between the two data sets were made, although I cannot measure how common this is.

I use these matched applications to rule out some potential, but unlikely, problems with the data. One concern is that bureaucrats receiving the scorecards found a way to manipulate the dates in the administrative data to improve their scorecards. This is unlikely since the administrative data was kept on a government server and could only be accessed by a few government contractors. I worked closely with this group since they were the ones transferring the data to me on a monthly basis and there are no suggestions that they were ever contacted by bureaucrats to alter the data on the server. Another concern is that applicants were pressured into saying that applications were processed faster than they actually were. This is unlikely since the interviews were by phone and there is no way for anyone from the land office to know what the applicant responded. Among the matched applications, the average processing time provided by the applicants in the survey was 67, while the same average time in the administrative data was 89. The average difference between the times was 22 in the control group and 20 in the offices receiving the scorecards. This suggests that there was no differential measurement bias between the treatment and control offices. The correlation between applications being processed within 45 working days in the administrative data and as measured by the processing time stated by the applicants was .25. The correlation in the log of processing times was .24. The low correlations suggests that the matching process suffers from a large number of false positive matches.

Appendix Table A15 shows the results, as estimated in Table 2, for matched applications only, using both the administrative data and the survey data on processing times. Restricting the applications to only matched observations reduces the precision of the estimates. However, the point estimates for the faction of applications processed within the 45 working day time limit, the outcome measure most important for the bureaucrats, are similar across the two data sets. This is reassuring that the overall effect is not driven by manipulations of the data by the bureaucrats.

B.2.5 Comparing bribe data with Transparency International National Household Survey Data

Bribes are notoriously difficult to measure precisely. Fortunately, bribe payments for the purpose of applications for changing government land records has was measured independently by Transparency International Bangladesh as part of their nationally representative National Household Survey (Transparency International Bangladesh, 2016). The survey took place in 2015 and asked about bribes paid between November 2014 and October 2015 and surveyed a nationally representative sample of households. The survey was done in person potentially allowing surveyors to build more rapport with the respondents. Since the survey was also done in a nationally representative sample of households in Bangladesh and for a different time period potentially causing the two estimates to be different for reasons other than survey methodology and statistical variation. 605 of the households had made applications for land record changes, among these 57% reported having paid a bribe. Appendix Figure A8 shows that the bribes reported in the Transparency International survey are on average higher than the bribes reported in the scorecard experiment phone survey, but that the difference shrinks substantially when excluding respondents reporting zero bribes. This could be due to that fewer respondents were comfortable to discuss bribe payments over the phone. The typical bribe payments reported in the scorecard experiment survey are slightly higher than the average payment reported in the Transparency International Bangladesh survey but lower than the average non-zero response. Overall, it is reassuring for the bribe results to have external validity within Bangladesh that the two different measures are not too different despite using different methodologies, covering different areas and being done for different time periods.

C Additional Empirical Analysis

C.1 External validity of results

The experiment was conducted at a large scale, with more than half of Bangladesh's land offices taking part in the experiment. The large scale of the experiment makes it plausible that the results are externally valid within Bangladesh (Muralidharan and Niehaus, 2017). The experiment included 59% of Bangladesh's land offices, covering an area with a population of approximately 95 million people. The experiment also spans a time frame of 16 months, reducing the concern for novelty effects (Jayaraman, Ray, and De Véricourt, 2016), and differences in effects across time periods (Rosenzweig and Udry, 2020). Moreover, the intervention was implemented with the Government of Bangladesh, the same organization that may eventually scale-up the policy. However, the scorecards were designed in a collaboration between a government agency and the author, and produced and distributed by Innovations for Poverty Action, a non-profit research organization. Hence, one should be cautious when extrapolating the results from the experiment to a potential scale-up by the Government itself (Bold, Kimenyi, Mwabu, Ng'ang'a, and Sandefur, 2018). Potential general equilibrium effects within the civil service are discussed in Section 6.3.

C.1.1 Geographic external validity within Bangladesh

The 311 land offices included in the experiment are the land offices that were actively using the e-governance system by the end of March 2019. Hence, it would be a problem for external validity if offices receiving the scorecard at different times had systematically different effects from the scorecards.

In Appendix Table A16, I test if this is the case for the effect on processing times by interacting the treatment with the date the land office started its first application using the e-governance system. Columns (1) and (2) show that the coefficients on the interaction term are close to zero for both the main outcome variables, the faction of applications processed within the time limit and the overall processing time. In Column (3) the coefficient on the interaction term for the ICW index is -0.005 (S.E. = 0.009). The point estimate is close to zero and my preferred interpretation is that the size of the effect from the scorecards does not vary with the installation date. However, one could interpret the point estimates as a decline in the effect of the scorecards for each month later that the e-governance system was installed. Using a linear prediction, the expected effect of the scorecards for the office in the experiment that had the latest installation date, is expected to be 0.080 standard deviations. The expected effect for the last offices to have the e-governance system installed in Bangladesh, in September 2019, is 0.057 standard deviations. Therefore, although this predicted effect is smaller for offices that had the e-governance system installed later, the scorecards are predicted to have a positive effect for all land offices in Bangladesh where the e-governance system is installed.

C.2 Potential bias from applicant survey and information intervention

Another potential threat to external validity is that the information intervention, or more generally the applicant survey, may have affected the behavior of bureaucrats and applicants. Since the information intervention and the applicant survey was carried out both in offices receiving scorecards and in control offices, it is unlikely that such effects have biased the results. To completely rule out the possibility of such bias, in Appendix Table A17, I conduct the main analysis for processing times from Table 2 using only applications that could not have been affected by the survey. I restrict the sample to applications from offices that were never surveyed and applications that were made more than 45 working days before the start of the survey in offices that were eventually surveyed. All of the estimates in Appendix Table A17 are very close to the estimates found in Table 2, ruling out any meaningful bias in the main estimates stemming from interactions between the applicant survey and the scorecards.

C.3 Unintended consequences of the scorecards

A common implementation problem of quantitative performance measures is that they lead to unintended and sometimes welfare reducing consequences (Banerjee et al., 2008; Rasul and Rogger, 2018). Below I briefly discuss several potential unintended consequences the performance scorecards could have led to. I do not find evidence for any substantial unintended effect except the effect on bribe payments taken by bureaucrats over-performing at baseline, which I discuss at length in the paper.

C.3.1 Improving indicators without improving service delivery

One potential concern is that land offices may reduce the number of applications, either by refusing to serve some applicants or by processing some applications using the paper-based system and not fully implement the e-governance system. With a smaller number of applications, it may be easier to reach a higher performance. Anticipating this problem the scorecards measure performance using the absolute number of applications processed within the time limit and not just the ratio. However, the number of applications pending beyond the time limit would still be easier to keep down with fewer applications. Another potential problem could be that bureaucrats only received applications which they knew would be easier to process. The size of the land for which the land record change is being made is positively associated with the processing time. Therefore, if bureaucrats indented to avoid accepting complex applications we would also see a decrease in the average land size.

Appendix Table A18 Column (1) shows that the scorecards did not substantially affect the number of applications received in the e-governance system. Column (2) shows that the scorecards did not decrease the number of scorecards more in the offices that were under-performing at baseline and had the largest improvement in processing times. Column (3) shows that the scorecards did not substantially affect the average land size among applications received, and Column (4) shows no evidence of a heterogeneous effect on land size among received applications. Overall the results are consistent with the bureaucrats not altering the applications received in response to the scorecards.

C.3.2 Quality of decision making

Another potential concern is that the quality of the decisions made by the bureaucrats was reduced by the scorecards. The main decision the bureaucrat makes with regards to the application is whether to reject or accept it. It is possible that when the bureaucrat spends less time on each application, more acceptances or more rejections are made depending on what the quickest action is to dispose of the application. Column (1) of Appendix Table A19 shows that the scorecards did not change the fraction of applications accepted. However, it is still possible that the quality of the decision was worse, meaning that more applications that should have been rejected were accepted and that more applications that should have been rejected were accepted. If an application is wrongfully rejected, applicants typically reapply in the same office. Therefore, the fraction of applicants reapplying after having been previously rejected can be used as an indicator for the fraction of incorrect rejections. Column (2) of Appendix Table A19 shows that the fraction of applicants stating that they were reapplying, after having been previously rejected for the same application, was not substantially increased by the scorecards. The results in Columns (1) and (2) together suggest that the scorecards did not lead to a decrease in the quality of decision making.

C.4 Effects of information intervention on bribes

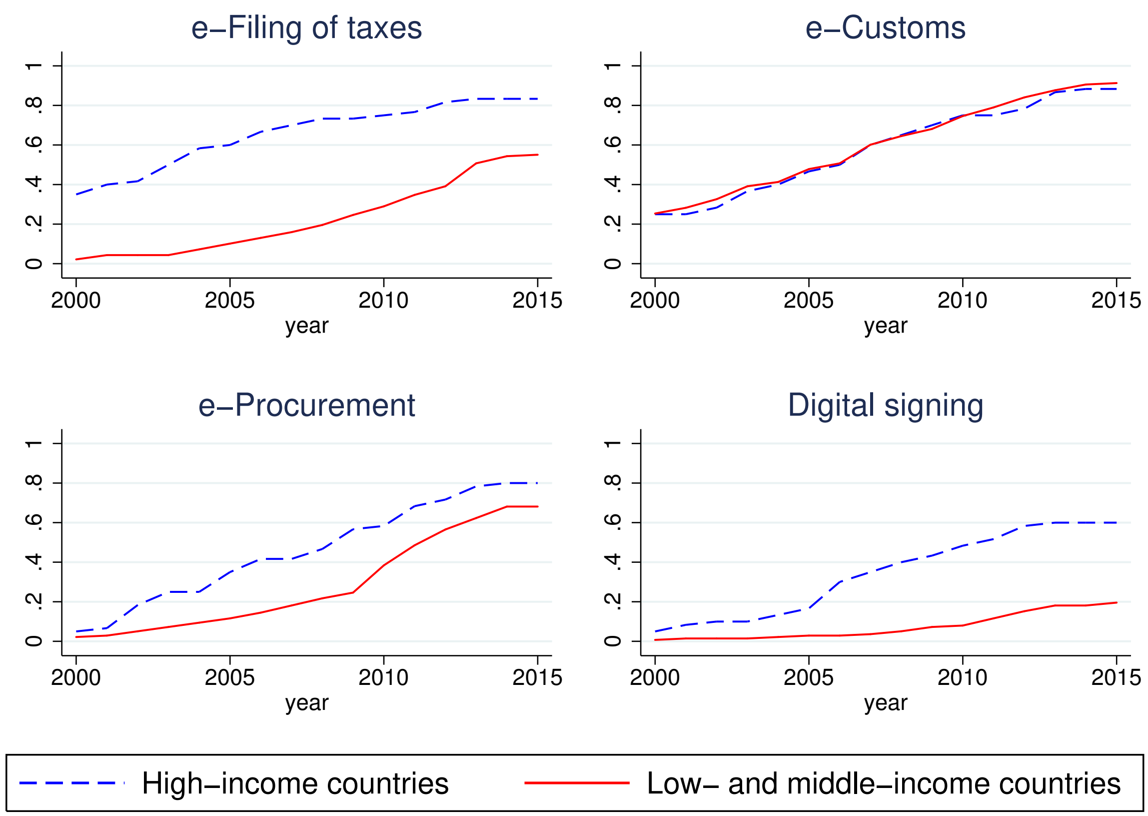
Appendix Table A4 shows that the information treatment did not have an effect on bribes, neither by itself nor in combination with the scorecards. Column (1) and (4) show that the information treatment by itself did not have a substantial impact on reported bribes or estimates of typical bribe payments. Columns (2) and (5) show that even together with the information treatment, the scorecards did not reduce bribes, neither for reported bribes nor for typical bribes. Columns (3) and (6) show that even among under-performing land offices, that improved the processing times the most, bribes did not decrease, neither with nor without the information treatment. Columns (3) and (6) also show that the positive effect of the scorecards on bribes, among offices over-performing at baseline, is similar across for applicants receiving the information intervention and the applicants not receiving the information intervention.

C.5 Effects of scorecard on applicant satisfaction

Appendix Table A20 shows the estimated effects of the scorecards on applicant satisfaction. Satisfaction was measured in the follow-up phone survey by asking applicants "Overall, how satisfied are you with the processing of your application?". The respondent could answer the question on a five-point scale ranging from very satisfied to not satisfied at all. The response was then transformed into standard deviations from the control group mean and used as an outcome variables in regression Equation 1 and 2. Column (1) of Appendix Table A20 shows the overall effect on satisfaction which is negative but small and not statistically significant. Column (2) splits up this effect between offices that were under-performing and offices over-performing at baseline. The negative effect is driven by offices that were over-performing at baseline which is consistent with the observation that the scorecards increased bribe payments increased in these offices. Furthermore, despite that the scorecards were successful in reducing processing for offices under-performing at baseline, the effect on the satisfaction stated by applicants in these offices is close to zero. Overall the results are consistent with a low valuation of faster processing times by the applicants but given the imprecise results it is hard to draw any definitive conclusion from the null result.

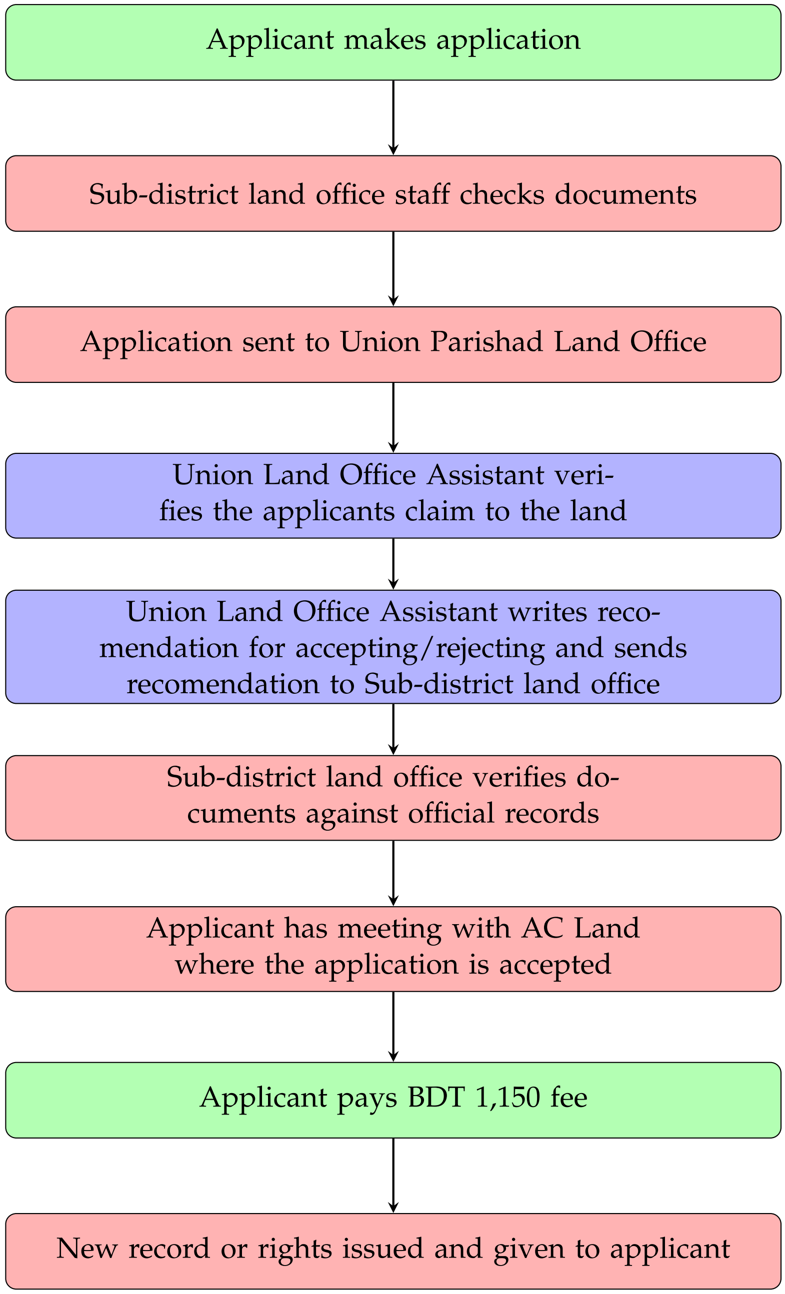
D Additional Tables and Figures

Figure A1: Digital Government Capacity by Income Group



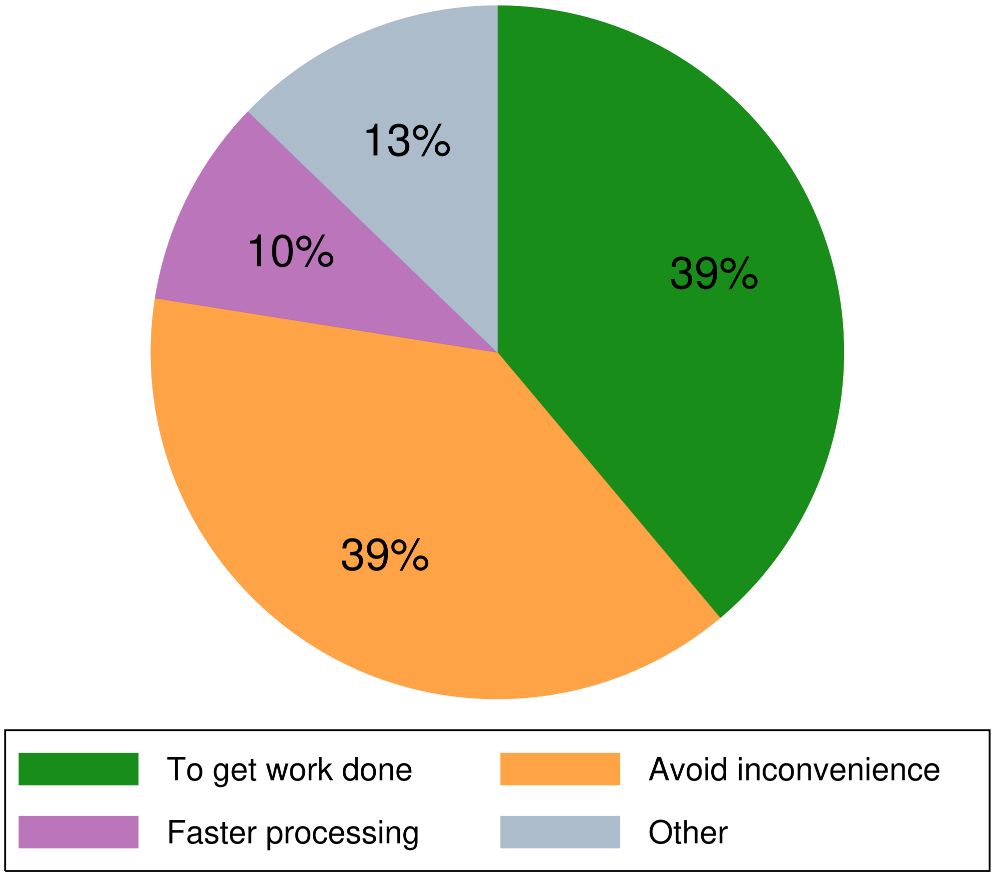
This figure shows what fraction of countries have at least a partially implemented e-governance system for four common interactions between the government and citizens or firms. The first is an e-governance system for filing taxes. The second is an e-governance system for clearing customs. The third is an e-governance system for public procurement. The fourth is legislation enabling digital signatures. Data is from the World Bank's *Public Financial Management Systems And E-Services Global Dataset* updated August 2017. Discussed in Section 2.

Figure A2: Application process for successful application



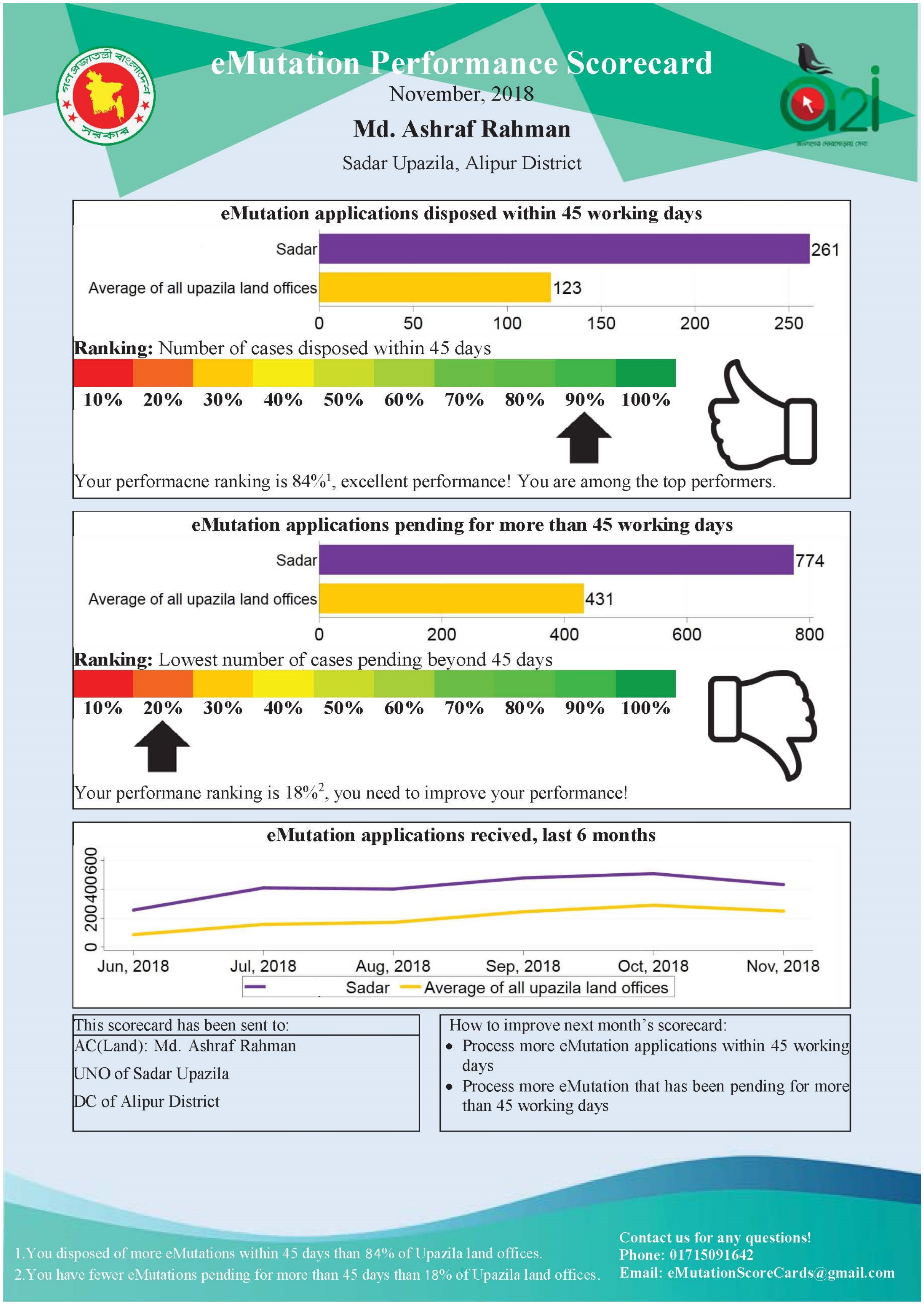
This figure provides a visual overview of the process of getting an application for a land record change approved. Green boxes represents actions by the applicant. Red boxes represents actions by the sub-district (*Upazila*) land office. Blue boxes represents actions by the local (*Union Parishad*) land office.

Figure A3: Stated reasons for bribe payments



This figure shows the reason stated by the applicants for paying bribes. The responses are weighted by the amount of the bribe. Therefore, the percentages should be interpreted as what percentage of the total bribe amounts were paid for what reason. The question was open-ended and was coded into response categories. Discussed in Sections 2.1 and 6.

Figure A4: Example of performance scorecard



This is an example of a performance scorecard in English. The ACL name and land office name are changed to preserve the anonymity of the civil servant. A version in Bengali was also included.

Figure A5: Example List of Offices Performances

List of eMutation Performance

November, 2019

The performance scorecard is sent monthly and evaluates a land office in terms of the number of eMutation applications disposed within 45 working days and the number of applications pending for more than 45 working days. A ranking from 0%-100% is then calculated for the number of applications disposed within 45 working days (disposed ranking) and the number of applications pending for more than 45 working days (pending ranking). A higher ranking percentage reflects a better performance.

We are now piloting sending out a list of all Upazilas' performance every month. Your Upazila has been selected to be part of this pilot project. The list below has been sent to 77 Upazilas.

Land offices can improve their performance scorecards by:

- Processing more eMutation applications within 45 working days
- Processing more eMutation applications that has been pending for more than 45 working days

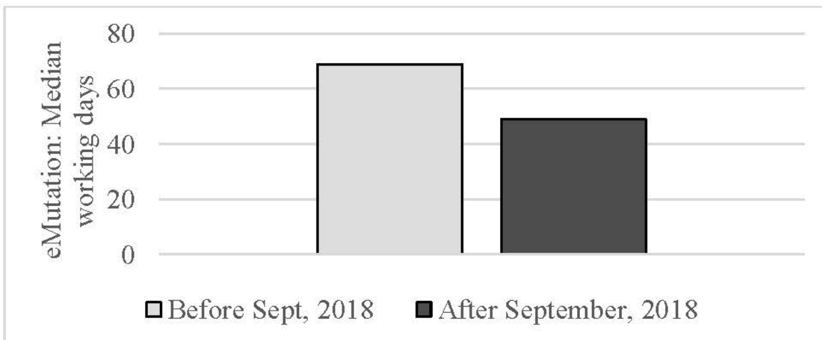
Upazila Land Office	District	Disposed ranking	Pending Ranking
[Name of upazila]	[Name of district]	43%	20%
[Name of upazila]	[Name of district]	48%	39%
[Name of upazila]	[Name of district]	51%	57%
[Name of upazila]	[Name of district]	17%	16%
[Name of upazila]	[Name of district]	18%	66%
[Name of upazila]	[Name of district]	52%	90%
[Name of upazila]	[Name of district]	45%	74%
[Name of upazila]	[Name of district]	98%	25%
[Name of upazila]	[Name of district]	28%	100%
[Name of upazila]	[Name of district]	97%	44%
[Name of upazila]	[Name of district]	5%	26%
[Name of upazila]	[Name of district]	26%	57%
[Name of upazila]	[Name of district]	69%	74%
[Name of upazila]	[Name of district]	56%	97%
[Name of upazila]	[Name of district]	50%	74%
[Name of upazila]	[Name of district]	56%	71%
[Name of upazila]	[Name of district]	26%	33%
[Name of upazila]	[Name of district]	22%	83%
[Name of upazila]	[Name of district]	95%	48%
[Name of upazila]	[Name of district]	44%	79%
[Name of upazila]	[Name of district]	52%	41%
[Name of upazila]	[Name of district]	18%	3%
[Name of upazila]	[Name of district]	80%	65%
[Name of upazila]	[Name of district]	81%	4%
[Name of upazila]	[Name of district]	97%	8%
[Name of upazila]	[Name of district]	99%	4%
[Name of upazila]	[Name of district]	5%	3%
[Name of upazila]	[Name of district]	96%	2%
[Name of upazila]	[Name of district]	81%	32%
[Name of upazila]	[Name of district]	93%	32%
[Name of upazila]	[Name of district]	64%	50%

This is an example of the first page of a list of a peer performance list, the full list contain two pages. The office and district names have been removed to preserve the anonymity of the civil servants.

Figure A6: Information pamphlet given in information intervention

Information for applicants
Land Record eMutation

Over the past 6 months the Government of Bangladesh have taken several steps to reduce the time it takes to process a Land Record Mutation. Before a typical Land Record eMutation took 69 working days (more than 3 months) now a typical Land Record eMutation takes 49 working days (a little more than 2 months).



You can apply for a Land Record eMutation by visiting the Upazila Land Office or from any computer connected to the internet (http://training.land.gov.bd/mutation/application).

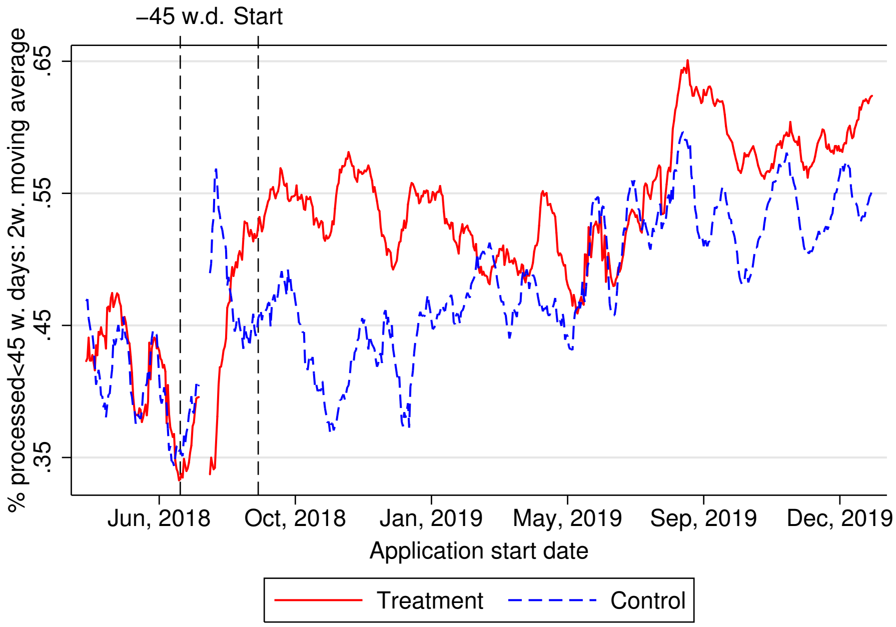
Steps of Land Record eMutation and timeline:

1. Make application online or in Upazila Land Office
2. Upazila Land Office will check the application and send it to Union Land Office
3. Union Land Office Assistant will make visit to land and write report to Upazila Land Office
4. Upazila Land Office will read report and call you for hearing via text message
5. You will attend hearing (according to text message)
6. Pay fee of 1150 taka and receive your Khatian

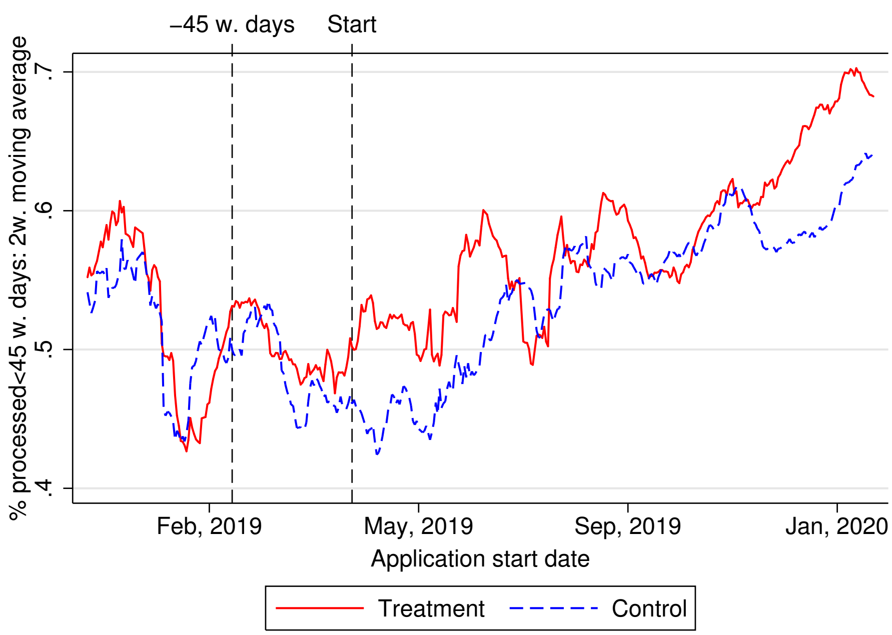
This information sheet was prepared by Innovations for Poverty Action in collaboration with a2i and the Land Reforms Board of Bangladesh.
Contact phone number: [redacted]

English translation of the information pamphlet given to applicants in the information intervention. The pamphlet shows the median application time for processed applications made before September 2018 and applications made after September 2018 in the 112 offices where the interviews took place. The data is as of February 2019. The same pamphlet was given to applicants in both treatment and control offices.

Figure A7: Fraction of applications processed within 45 working days



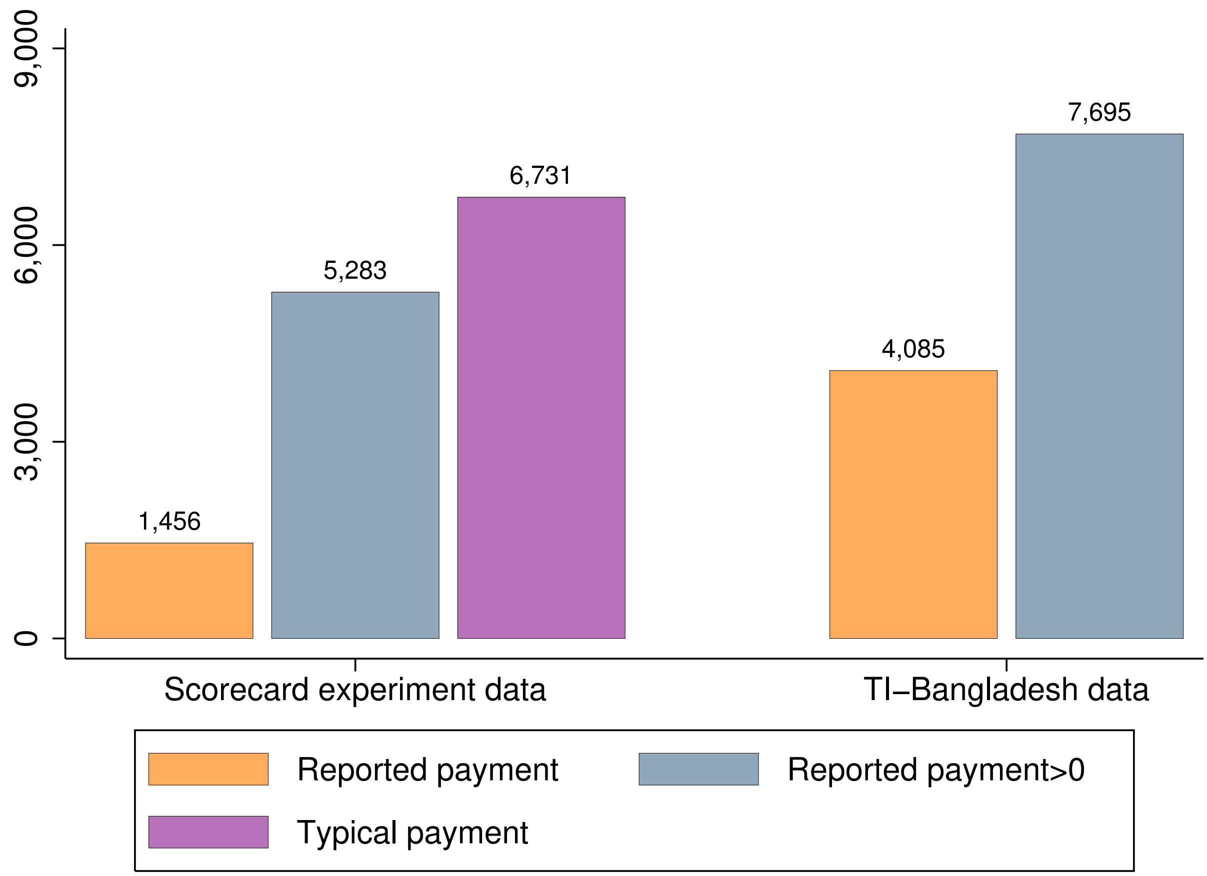
(a) First randomization wave



(b) Second randomization wave

This figure shows the two-week moving averages of the fraction of applications processed within the 45 working day limit in the treatment and control groups. Sub-figure A8a shows data from the first randomization wave Sub-figure A8b shows data from the second randomization wave. The effect of the treatment can in principle have started for applications started 45 working days before the first scorecard was sent (first vertical dashed line) but only application made after the first scorecard was sent (second vertical dashed line) were fully treated. The gap in the Sub-Figure A8a time-line is due to a server error that caused the e-governance system to temporarily shut down in late July 2018. Discussed in Section 4.1.1.

Figure A8: Estimated average bribe payments compared to estimates from Transparency International Bangladesh



This figure shows the average bribe payments reported in the phone survey conducted to evaluate the scorecard experiment and in an independent survey by Transparency International Bangladesh. The first bar shows the average value of bribe payments reported by the applicant in the scorecard experiment phone survey, 73% of the applicants reported having paid no bribes. The second bar shows the average value of bribe payments reported by applicants reporting having paid some bribe in the scorecard experiment phone survey. The third bar shows the average value of an estimated "typical bribe payment by a person like yourself" reported in the scorecard experiment phone survey, 27% of the applicants responding to this question reported that a typical applicant paid no bribes. The fourth bar shows the average value of bribe payments reported by the applicants in the Transparency International Bangladesh survey, 57% of the respondents reported having paid no bribe for their land record change. The fifth bar shows the average value of bribe payments reported by respondents reporting having paid some bribe in the Transparency International Bangladesh survey. All variables are winsorized at the 99th percentile. Observations in the three first bars are are inversely weighted by the number of observations in that land office. Discussed in Appendix Section B.2.5.

Table A1: Balance of randomization: Administrative data

	(1) Control		(2) Scorecard		(3) T-test (1)-(2)
Variable	N/[Clusters]	Mean/SD	N/[Clusters]	Mean/SD	Difference/SE
<45 w. days	56,686	0.422	57,028	0.444	-0.022
	[146]	(0.49)	[146]	(0.50)	(0.08)
IHS(w. days)	56,686	4.744	57,028	4.734	0.010
	[146]	(1.02)	[146]	(1.11)	(0.199)
Process time (w. days)	56,686	88.274	57,028	96.456	-8.181
	[146]	(78.75)	[146]	(95.96)	(20.2)
Approved before experiment start	38,728	0.738	36,272	0.751	-0.013
	[141]	(0.44)	[136]	(0.43)	(0.06)

This table shows the balance of randomization for treatment and control offices using administrative data from 45 working days before the first scorecard was sent (this date is different for randomization wave 1 and randomization wave 2 offices). Due to this restriction only 292 of the 311 offices are part of the balance of randomization data. Applications not processed by the first scorecard had the processing time imputed using the procedure described in Section 2.5.1. Data on approvals are as per the start of the treatment. P-value for F-test of joint orthogonality: 0.83. Observations are uniformly weighted. When using the weighted regression specification from Equation 1 on this data the effect of the treatment is not statistically significantly different from zero for any of the outcome variables. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 2.6.

Table A2: Balance of randomization: Survey data

	(1) Control		(2) Scorecard		(3) T-test (1)-(2)
Variable	N/[Clusters]	Mean/SD	N/[Clusters]	Mean/SD	Difference/SE
Applicant age	1,463	47.33	1,440	47.37	-0.04
	[56]	(13.83)	[56]	(13.20)	(0.62)
Female	1,498	0.07	1,520	0.06	0.01
	[56]	(0.25)	[56]	(0.24)	(0.01)
Monthly income (BDT)	1,407	28,505	1,384	32,568	-4,063
	[56]	(92,587)	[56]	(133,391)	(5,832)
App. status: Applying	1,498	0.24	1,520	0.20	0.03
	[56]	(0.42)	[56]	(0.40)	(0.04)
Ongoing	1,498	0.60	1,520	0.61	-0.004
	[56]	(0.49)	[56]	(0.49)	(0.04)
Rejected	1,498	0.002	1,520	0.005	-0.003
	[56]	(0.04)	[56]	(0.07)	(0.003)
Approved	1,498	0.07	1,520	0.07	0.01
	[56]	(0.26)	[56]	(0.25)	(0.02)
Land Value (BDT 100k)	1,418	18.21	1,382	22.09	-3.87
	[56]	(44.30)	[56]	(53.42)	(2.85)
Land Size (Decimal)	1,455	25.42	1,437	25.66	-0.24
	[56]	(43.44)	[56]	(52.61)	(2.91)

All data comes from the in-person survey of applicants, which was conducted before the conclusion of the processing of the application, but after the start of the scorecards. USD/BDT≈84.3. 1 decimal = 1/100 acre. P-value for F-test of joint orthogonality: 0.89. Observations are uniformly weighted. When using the weighted regression specification from Equation 1 on this data, the effect of the scorecards is not significant at the 5% for any of the outcome variables, and significant at the 10% level only for land value. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 2.6.

Table A3: Testing prediction from monopolistic price discrimination model

	Reported payment				
	(1)	(2)	(3)	(4)	(5)
Scorecard	656.4***	332.7		903.3***	
	(214.6)	(259.8)		(319.2)	
Scorecard x Overperform			635.2***		
			(227.2)		
Scorecard x Underperform			686.2*		
			(399.1)		
Overperform baseline			9.139		12.19
			(322.2)		(317.6)
Information treatment				-111.2	-110.0
				(239.1)	(238.2)
Scorecard x Information				-430.4	
				(431.3)	
Info x Scorecard x Underperform					499.7
					(512.0)
No info x Scorecard x Underperform					946.6*
					(495.2)
Info x Scorecard x Overperform					452.2
					(295.5)
No info x Scorecard x Overperform					875.5**
					(385.7)
Start month FE	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes
Sample	Appr. < 25	Appr. > 25	Appr. < 25	Appr. < 25	Appr. < 25
Observations	672	1,447	672	672	672
Clusters	109	111	109	109	109

This table shows the effect of the scorecard and information treatments on bribes made for application processing. Column (1) and (3)-(5) use data only from applications that were processed within 25 working days while Column (2) uses data from applications that were processed for more than 25 working days. USD/BDT≈84.3 Continuous variables winsorized at the 99th percentile. Standard errors are clustered at the land office level. ***p<0.01; **p<0.05; *p<0.1. Discussed in Appendix Section A.2.

Table A4: Effect of information treatment and scorecards on bribes for application processing

	Reported payment			Typical payment		
	(1)	(2)	(3)	(4)	(5)	(6)
Information treatment	1.638	-10.88		252.1	-189.9	
	(146.5)	(187.0)		(495.4)	(749.7)	
Scorecard		261.0			644.9	
		(222.6)			(720.1)	
Scorecard x Information		8.499			812.3	
		(274.0)			(1098.8)	
Info x Scorecard x Overperform			674.2**			1761.7**
			(268.4)			(871.3)
No info x Scorecard x Overperform			603.9**			2747.7***
			(253.3)			(873.7)
Info x Scorecard x Underperform			-94.80			999.7
			(287.0)			(1239.5)
No info x Scorecard x Underperform			-17.53			-1440.1
			(315.7)			(875.8)
Overperform baseline			-819.9***			-1766.7*
			(287.9)			(924.2)
Start month FE	Yes	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,018	3,018	3,018	1,896	1,896	1,896
Clusters	570	112	112	544	112	112
Control mean	1,467	1,302		6,512	6,157	

This table shows the effect of the scorecard and information treatments on bribes made for application processing. Column (1)-(3) estimate the effects on reported bribes. Columns (4)-(6) estimate the effects on typical bribe payments. USD/BDT≈84.3 All outcome variables are winsorized at the 99th percentile. In Columns (1) and (4) standard errors are clustered at the land office by day level. In Columns (2)-(3) and (5)-(6) standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Appendix Section 5.1.

Table A5: Robustness of effect on processing time with respect to imputation technique

	(1) Mean	(2) Office mean	(3) Rand. obs.	(4) No impute
Scorecard	-0.125**	-0.119**	-0.123**	-0.120**
	(0.0593)	(0.0579)	(0.0582)	(0.0511)
Constant	4.417***	4.402***	4.406***	4.294***
	(0.0403)	(0.0389)	(0.0395)	(0.0353)
Start month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	1,050,924	1,050,924	1,050,924	972,589
Clusters	311	311	311	311

This table shows the robustness of the result in Column (2) of Table 2 with regards to the imputation procedure used to assign a processing time to applications that are not yet processed. Column (1) uses the mean of processing times for all applications that were processed after the number of days that the application that I am imputing the processing time for has been pending. Column (2) uses the mean of processing times for applications in that land office that were processed after the number of days that the application that I am imputing the processing time for has been pending. Column (3) uses a randomly selected processing time of the processing times that are larger than the number of days that the application that I am imputing the processing time for has been pending. Column (4) drops all applications that are not yet processed from the sample. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 4.1.

Table A6: Robustness of effect on processing time with respect to functional form assumption

	ln(Days)	Working Days		
	(1)	(2)	(3)	(4)
Scorecard	-0.123**	-6.492	-0.101	-0.0980*
	(0.0587)	(4.033)	(0.0646)	(0.0585)
Start month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	1,048,876	1,050,924	1,050,924	1,050,924
Clusters	311	311	311	311
Specification	OLS	OLS	Poisson	Neg. Binomial

This table shows the robustness of the effect of the scorecards on the processing time to different assumptions regarding the functional form of the relationship between the processing time and the scorecards. Column (1) uses the natural logarithm transformation treating observations with the value zero as missing. Column (2) uses the untransformed number of working days. Column (3) shows the results of a Poisson regression. Column (4) shows the results of a negative binomial regression. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 4.1.

Table A7: Estimating the effect on processing time with alternative specifications

	ICW Index				
Panel A. Overall effect	(1)	(2)	(3)	(4)	(5)
Scorecard	0.081	0.125*	0.126**	0.091	0.129**
	(0.091)	(0.069)	(0.059)	(0.078)	(0.055)
Panel B. Heterogeneous effects					
Scorecard x Overperform baseline	-0.016	-0.002	0.002	0.023	0.020
	(0.115)	(0.089)	(0.078)	(0.097)	(0.075)
Scorecard x Underperform baseline	0.210*	0.283***	0.272***	0.190*	0.259***
	(0.121)	(0.093)	(0.087)	(0.113)	(0.082)
Overperform baseline	0.558***	0.580***	0.311***	0.497***	0.237**
	(0.108)	(0.087)	(0.090)	(0.131)	(0.107)
Start month FE	No	No	No	Yes	Yes
Stratum FE	No	No	No	Yes	Yes
Weighted by office	No	Yes	Yes	No	Yes
Baseline controls	No	No	Yes	No	Yes
Observations	1,050,924	1,050,924	1,050,924	1,050,924	1,050,924
Clusters	311	311	311	311	311

This table shows the robustness of the estimated effect of the scorecards on the IWC Index of being processed within the time limit and the IHS of processing time to different regression specifications. Panel A shows the estimates of the overall effect similar to the estimates in Table 2. Panel B shows the estimates of the heterogeneous effects, similar to the estimates in Table 6. In what follows I describe how the specifications differ from the specifications in those tables. Column (1) shows the estimate from an unweighted regression with no fixed effects. Column (2) shows the estimate from a regression with no fixed effects. Column (3) shows the estimate from a regression with no fixed effects, controlling for baseline measures of the number of applications processed within 45 working days, the number of applications pending beyond 45 working days, applications received and the fraction of applications processed within 45 working days. Column (4) shows the estimate from an unweighted regression. Column (5) shows the estimate from a regression controlling for baseline measures of the number of applications processed within 45 working days, the number of applications pending beyond 45 working days, applications received and the fraction of applications processed within 45 working days. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 4.1.

Table A8: Estimating the effect of Scorecards on office by month level outcomes

Panel A. Overall effect	(1) IHS Dis.$\leq$45	(2) IHS Pen. > 45	(3) Rank dis.	(4) Rank pen.	(5) ICW index
Scorecard	0.212*	-0.058	2.133	1.464	0.080
	(0.121)	(0.140)	(1.685)	(1.782)	(0.062)
Panel B. Heterogeneous effects					
Scorecard x Overperform	-0.039	0.280	-0.819	-2.524	-0.099
	(0.149)	(0.212)	(2.209)	(2.609)	(0.082)
Scorecard x Underperform baseline	0.499**	-0.441**	5.563**	5.899**	0.285***
	(0.195)	(0.180)	(2.587)	(2.356)	(0.091)
1st PS>median	0.537**	-0.286	7.685**	1.944	0.242*
	(0.236)	(0.264)	(3.425)	(3.401)	(0.127)
Month FE	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes
Observations	4,516	4,516	4,516	4,516	4,516
Clusters	311	311	311	311	311

This table shows the effect of the scorecards on office by month level outcomes. Panel A shows the estimates of the overall effect similar to the estimates in Table 2. Panel B shows the estimates of the heterogeneous effects, similar to the estimates in Table 6. Column (1) shows the effect on the IHS of the number of applications processed within 45 working days. Column (2) shows the effect on the IHS of the number of applications pending beyond 45 working days. Column (3) shows the effect on the percentile ranking in terms of the number of applications processed within 45 working days. Column (4) shows the effect on the percentile ranking in terms of the number of applications processed within 45 working days, a higher number of pending applications leads to a lower ranking. Column (5) shows the result on a ICW index created with the outcome variables of Tables (1)-(4). Standard errors are clustered at the land office level. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 4.1.

Table A9: Estimating the effect on bribes with alternative specifications

	Typical payments					Reported payments					To gov. off.
Panel A. Overall effect	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Scorecard	1,324*	1,277*	978	1,974**	566	384*	355*	288	564**	175	214
	(699)	(700)	(613)	(844)	(443)	(201)	(195)	(174)	(263)	(125)	(137)
Panel B. Heterogeneous effects											
Scorecard x Overperform	2,294**	2,034**	2,467***	3,055***	1,637***	741***	708***	680***	1,066**	454***	416**
	(931)	(877)	(845)	(1,070)	(583)	(235)	(215)	(229)	(426)	(163)	(165)
Scorecard x Underperform	435	557	-486	938	-480	67	29	-52	80	-66	55
	(1,011)	(1,062)	(883)	(1,309)	(640)	(312)	(307)	(245)	(364)	(168)	(197)
Overperform baseline	-1,314	-1,504	-2,010**	-1,945*	-1,616**	-515*	-700***	-778***	-660*	-637***	-585**
	(813)	(916)	(921)	(1,100)	(688)	(262)	(266)	(282)	(386)	(195)	(227)
Start month FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Stratum FE	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Weighted by office	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes
Winsorized	99 pctl.	99 pctl.	99 pctl.	No	95 pctl.	99 pctl.	99 pctl.	99 pctl.	No	95 pctl.	99 pctl.
Observations	1,896	1,896	1,896	1,896	1,896	3,018	3,018	3,018	3,018	3,018	3,018
Clusters	112	112	112	112	112	112	112	112	112	112	112

This table shows the robustness of the estimated effect of the scorecards on bribes. Panel A shows the estimates of the overall effect similar to the estimates in Columns (1) and (2) of Table 5. Panel B shows the estimates of the heterogeneous effects, similar to the estimates in Columns (1) and (2) of Table 7. Columns (1)-(5) show the effect on the estimate for how much a "normal person, like yourself" pays in bribes. Columns (6)-(10) show the effect on reported payments to government officials or agents beyond the official fee. In what follows I describe how the specifications differ from the specifications in Tables 5 and 7. Columns (1) and (6) show the estimate from an unweighted regression with no fixed effects. Columns (2) and (7) show the estimate from a regression with no fixed effects. Columns (3) and (8) show the estimate from an unweighted regression. Columns (4) and (9) show the estimate when not winsorizing the outcome variable. Columns (5) and (10) show the estimate when winsorizing the outcome variable at the 95th percentile. Column (11) shows the estimate of the effect on payments made directly to government officials, excluding payments made to agents. Standard errors are clustered at the land office level. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 4.4.

Table A10: Robustness of effect heterogeneity to alternative measures of baseline performance

	Processing time: ICW Index			Typical payments			Reported payments		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Scorecard x Overperform 3m baseline	0.0157			2097.1***			618.6***		
	(0.0802)			(712.1)			(213.8)		
Scorecard x Underperform 3m baseline	0.261***			125.9			-28.53		
	(0.0867)			(940.1)			(258.4)		
Treat x 4th quartile baseline		-0.0343			2335.1**			550.2**	
		(0.104)			(923.3)			(271.2)	
Treat x 3rd quartile baseline		0.0186			2425.3**			758.4**	
		(0.119)			(1182.4)			(365.2)	
Treat x 2nd quartile baseline		0.172			0.156			111.6	
		(0.122)			(967.7)			(299.4)	
Treat x 1st quartile baseline		0.353***			-234.2			-186.7	
		(0.118)			(1510.1)			(396.0)	
Treat x Baseline ranking			-0.00731**			41.23			16.07*
			(0.00305)			(29.05)			(8.840)
Scorecard			0.116**			1090.6*			299.8*
			(0.0573)			(602.5)			(171.4)
P-value sub-group diff.	0.04			0.10			0.05		
Start month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline performance control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,050,924	1,050,924	1,050,924	1,896	1,896	1,896	3,018	3,018	3,018
Clusters	311	311	311	112	112	112	112	112	112

This table shows the robustness of the results for the heterogeneity of the effect of the scorecards for different baseline performance measures. Columns (1)-(3) show the effects on the index of the two main processing time outcome variables used in Column (3) of Table 2. Columns (4)-(6) show the effect on the estimate for how much a "normal person, like yourself" pays in bribes. Columns (7)-(9) show the effects on the bribe payments reported by the applicant. Columns (1), (4), and (7) show the heterogeneity in the effect of the scorecards based on the office having an above- or below-median average ranking across the last three months of the baseline period. Columns (2), (5), and (8) show the heterogeneity based on the quartile of baseline ranking. Columns (3), (6), and (9) show the heterogeneity based on the continuous baseline ranking. USD/BDT≈84.3. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 4.5.

Table A11: Effect of information treatment and scorecards on expected processing time

	ln(Expected processing time)				
	(1)	(2)	(3)	(4)	(5)
Scorecard	-0.0886**		-0.0778		
	(0.0439)		(0.0511)		
Information treatment		-0.0373	-0.0262		
		(0.0276)	(0.0235)		
Scorecard x Information			-0.0211		
			(0.0439)		
Scorecard x Overperform				-0.0640	
				(0.0682)	
Scorecard x Underperform				-0.0944*	
				(0.0562)	
Info x Scorecard x Overperform					-0.0758
					(0.0666)
No info x Scorecard x Overperform					-0.0526
					(0.0784)
Info x Scorecard x Underperform					-0.133*
					(0.0681)
No info x Scorecard x Underperform					-0.0526
					(0.0549)
Overperform baseline				-0.137*	-0.138**
				(0.0691)	(0.0690)
Start month FE	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes
Observations	2,657	2,657	2,657	2,657	2,657
Clusters	112	528	112	112	112
Control mean	56	57	57		

This table shows the effect of the scorecard and information treatments on expected processing times at the time of the in-person interview. The outcomes variable is winsorized at the 99th percentile. In Column (1) standard errors are clustered at the land office by day level. In Columns (2) and (3) standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 6.4.

Table A12: Effect on bureaucrat transfers

	(1) Transfer	(2) Transfer	(3) Duration	(4) Duration	(5) No ACL	(6) No ACL
Scorecard	0.00210		0.533		-0.000236	
	(0.00554)		(0.671)		(0.0241)	
Scorecard x Overperform		0.00135		0.503		0.0103
		(0.00792)		(0.891)		(0.0335)
Scorecard x Underperform		0.00288		0.563		-0.0125
		(0.00821)		(1.037)		(0.0364)
Overperform baseline		0.00589		-0.202		-0.00531
		(0.00962)		(1.135)		(0.0346)
P-value: sub-group diff.		0.90		0.97		0.65
Month FE	Yes	Yes	No	No	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,516	4,516	304	304	4,516	4,516
Clusters	311	311	304	304	311	311
Control mean	0.07		12.22		0.13	
Control mean		0.07		12.19		0.12
Overperformers: control mean		0.07		12.26		0.14

This table shows the effect of the scorecards on transfers of ACLs. Columns (1) and (2) show the effects on the fraction of ACLs transferred away from the office in a particular office-month, using data for each office month after the start of the experiment until the last month of the experiment (March 2020). Columns (3) and (4) show the effect on the duration of the posting in months for the first bureaucrat to hold the position as ACL in each of the offices in the experiment. Columns (5) and (6) show the effect on not having any ACL in a particular office-month. The data is administrative data from the e-governance system. Standard errors are clustered at the land office level, except for in Columns (3) and (4), where heteroskedasticity robust standard errors are used. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section 6.4.2.

Table A13: Treatment effects on survey attrition

	Attrition			
	(1)	(2)	(3)	(4)
Scorecard treatment	0.0293*		0.0256	
	(0.0160)		(0.0194)	
Information treatment		0.00223	-0.00254	
		(0.0121)	(0.0174)	
Scorecard x Information			0.00708	
			(0.0233)	
Scorecard x Overperform baseline				0.0123
				(0.0214)
Scorecard x Underperform baseline				0.0476**
				(0.0240)
Overperform baseline				0.00461
				(0.0253)
Start month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	3,696	3,695	3,695	3,696
Clusters	112	112	112	112

This table shows the effect of the scorecards and information treatment on attrition from the survey. Attrition is measured from a survey being attempted at the time of the in-person survey until having a successful follow-up survey by phone. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Section B.2.2.

Table A14: Lower Lee bounds for the effects on bribe payments

	Reported payment	
	(1)	(2)
Scorecard treatment	204.88	
	(179.72)	
Scorecard x Overperform baseline		638.51***
		(224.25)
Scorecard x Underperform baseline		-177.14
		(255.68)
Overperform baseline		-820.76***
		(287)
Start month FE	Yes	Yes
Stratum FE	Yes	Yes
Weighted by office	Yes	Yes
Observations	3,075	3,076
Clusters	112	112
Control mean	1,278	
Overperformers: Control mean		916
Underperformers: Control mean		1,616

This table shows the lower Lee bounds (Lee, 2009) for the estimates of the effects of the scorecards on bribe payments shown in Column (1) of Table 5 and Column (1) of Table 7. A number, equal to the differential attrition rate, of randomly selected observations from the treatment group are added back into the data and assigned a value of zero for bribe payments. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. The table is discussed in Appendix Section B.2.3.

Table A15: Comparison of results using administrative and survey data

	(1)	(2)	(3)	(4)
	<45 w. days	<45 w. days	ln(w. days)	ln(w. days)
Scorecard	0.0449	0.0604	-0.141	-0.0250
	(0.0699)	(0.0509)	(0.168)	(0.0900)
Start month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	1,367	1,367	1,367	1,367
Clusters	108	108	108	108
Control mean	0.44	0.51	102.09	69.33

This table compares the effects estimated using the administrative and survey data. The regression specification is the same as in Table 2. All observations are applications matched between the survey and administrative data. Columns (1) and (3) show the results estimated using the matched administrative data. Columns (2) and (4) show the results estimated using the matched survey data. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Appendix Section B.2.4.

Table A16: Effect of scorecards by date e-governance system installed

	(1) <45 w. days	(2) IHS(w. days)	(3) ICW index
Scorecard	0.0568**	-0.117**	0.122**
	(0.0275)	(0.0596)	(0.0593)
Treat x installation date	-0.00303	0.00197	-0.00467
	(0.00409)	(0.00895)	(0.00887)
E-governance installation date	-0.0109	0.0253	-0.0246
	(0.00812)	(0.0179)	(0.0177)
Start month FE	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes
Observations	1,050,924	1,050,924	1,050,924
Clusters	311	311	311
Control mean	0.56	65.64	
Fraction imputed		0.06	
Fraction zero		0.003	

This table shows differences in the effects of the scorecards between offices that had the e-governance system installed during different time periods. The e-governance installation date is the date the first application was made using the e-governance system made in that office. The unit of the installation date variable is months and the variable is measured relative to the weighted mean of installation dates among the offices in the sample. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Appendix Section C.1.1. Discussed in Appendix Section C.1.

Table A17: Effect of scorecards, excluding applications potentially affected by applicant survey

	(1) <45 w. days	(2) IHS(w. days)	(3) ICW index
Scorecard	0.0683**	-0.168**	0.140**
	(0.0293)	(0.0668)	(0.0593)
Start month FE	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes
Observations	545,742	545,742	545,742
Clusters	310	310	310
Control mean	0.52	79.75	-0.00
Fraction imputed		0.05	
Fraction zero		0.003	

This table shows the results from Table 2 when restricting the sample to applications that were made either in offices where the applicant survey did not take place, or made 1 month or more before the start of the applicant survey. Hence these results are highly unlikely to have been affected by the survey activities. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Appendix Section C.1.

Table A18: Effect on the number of applications received and land size of application received

VARIABLES	(1) IHS apps. rec.	(2) IHS apps. rec.	(3) ln(Land size)	(4) ln(Land size)
Scorecard	-0.0517		0.0229	
	(0.0737)		(0.0674)	
Scorecard x Overperform baseline		-0.0577		0.0243
		(0.107)		(0.0946)
Scorecard x Underperform baseline		-0.0431		0.0214
		(0.107)		(0.0979)
Underperform baseline		-0.157		-0.000536
		(0.124)		(0.112)
Observations	311	311	1,042,987	1,042,987
Stratum FE	Yes	Yes	Yes	Yes
Start month FE			Yes	Yes
Weighted by office			Yes	Yes
Clusters			311	311

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

This table shows the effect of the scorecards on the number of applications received and the land size of those applications. In Columns (1) and (2) observations are at the office level. In Columns (3) and (4) the observations are at the application level. Data contains all applications made between 1 month before the start of the experiment started and 45 working days before the experiment ended (13 Aug 2018 - 20 Jan 2020). Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Appendix Section C.3.

Table A19: Effects on rejections

	(1)	(2)
	Prev. rejected	Prev. rejected
Scorecard treatment	-0.00810	0.0215
	(0.0196)	(0.0203)
Start month FE	Yes	Yes
Stratum FE	Yes	Yes
Weighted by office	Yes	Yes
Observations	1,050,924	3,215
Clusters	311	112
Control mean	0.29	0.06

This table shows the effect of the scorecards on rejections of applications for land record changes. Column (1) shows the effect of the scorecards on the fraction of applications rejected in the administrative data. Column (2) shows the effect on the fraction of applicants surveyed who was returning after having had their application rejected, which is a proxy for an incorrect rejection. Standard errors are clustered at the office level. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Appendix Section C.3.2.

Table A20: Effect on applicant satisfaction

	(1) Satisfaction	(2) Satisfaction
Scorecard	-0.0477	
	(0.0612)	
Scorecard x Overperform		-0.117
		(0.0827)
Scorecard x Underperform		0.00746
		(0.0897)
Overperform baseline		0.200**
		(0.0975)
P-value sub-group diff.		0.32
Start month FE	Yes	Yes
Stratum FE	Yes	Yes
Weighted by office	Yes	Yes
Observations	3,018	3,018
Clusters	112	112

This table shows the effect of the scorecards on applicants stated satisfaction transformed from a five-point scale into standard deviations away from the control group mean. Standard errors are clustered at the land office level. Observations are inversely weighted by the number of applications in that land office. ***p<0.01; **p<0.05; *p<0.1. Discussed in Appendix Section C.5.