Information Systems, Service Delivery, and Corruption: Evidence From the Bangladesh Civil Service

Martin Mattsson*

26th March 2022

Abstract

Government bureaucracies in low- and middle-income countries often suffer from both corruption and slow public-service delivery. Can an information system – providing information about delays to the responsible bureaucrats and their supervisors – improve delivery speed? Paying bribes for faster service delivery is common, but does improving the average delivery speed 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 on-time service delivery by 11% but do not reduce bribes. Instead, the scorecards *increase* bribes for high-performing bureaucrats. The results are inconsistent with a causal link between the average speed of service delivery and corruption, and they contradict several existing theories of this relationship. I propose a model where bureaucrats' reputational concerns or shame constrain corruption. When bureaucrats' reputations improve through positive performance feedback, this constraint is relaxed, and bribes increase.

JEL Codes: D02, D73, H83, O10

*National University of Singapore, martin.mattsson@nus.edu.sg. I would like to thank Nathan Barker, Gaurav Chiplunkar, Anir Chowdhury, Andrew Foster, Eduardo Fraga, Sahana Ghosh, Marina Halac, Ashraful Haque, Enamul Haque, Daniel Keniston, Mushtaq Khan, Ro'ee Levy, Imran Matin, Mushfiq Mobarak, Farria Naeem, Rohini Pande, Dina Pomeranz, Mark Rosenzweig, Nick Ryan, Yogita Shamdasani, Jeff Weaver, Jaya Wen, Fabrizio Zilibotti, and numerous seminar participants for helpful comments and suggestions. I also thank Muhammad Bin Khalid, Mahzabin Khan, and Ashraf Mian for excellent research assistance, IPA Bangladesh for outstanding research support. A randomized-controlled-trial registry and pre-analysis plan are available at: www.socialscienceregistry.org/trials/3232. This project was approved by Yale University IRB (Protocol ID 2000021565). 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

A government's capacity to implement its policies, secure property rights, and provide basic public services is paramount for economic development. To have this capacity, states need functioning bureaucracies with government officials motivated to carry out these tasks. 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 or political resistance (Banerjee et al., 2008; Dhaliwal and Hanna, 2017). Another approach is to improve systems that measure bureaucrats' performance. This may improve incentives by allowing supervisors to let job performance determine postings and promotions, a strong motivator for civil servants (Khan et al., 2019; Bertrand et al., 2020). Regular performance feedback may also make the performance more salient to the bureaucrats themselves, potentially leveraging government officials' intrinsic motivation (Prendergast, 2007; Banuri and Keefer, 2016).

This paper studies an information system designed to improve processing times of applications for changes to government land records in Bangladesh. Slow public-service delivery is a substaintial problem in Bangladesh. Among households that paid a bribe for a public service, 23% stated that "timely service" was one of the reasons for the bribe (Transparency International Bangladesh, 2018). This suggests that slow service delivery on average may cause corruption, as some citizens and firms pay bribes to avoid having to wait for the service. This is a common hypothesis in the literature (e.g. Myrdal, 1968; Rose-Ackerman, 1978; Kaufmann and Wei, 1999), but the causal relationship has not been established empirically.

In an experiment with the Bangladesh Civil Service, I provide information on junior bureaucrats' performance using monthly scorecards sent to the bureaucrats and their supervisors. The scorecards, designed to reduce delays in processing applications for land-record changes, are based on data from an e-governance system. Two performance indicators appear on the scorecards: the number of applications processed within a time limit of 45 working days and the number of applications pending beyond that limit. The scorecards also show bureaucrats' performance on these indicators relative to all other bureaucrats in the experiment. The intervention is randomized at the level of the subdistrict land office and involves 311 land offices (60% of all land offices in Bangladesh), serving a population of approximately 97 million people.

The scorecards improve processing times. Using administrative data from more than a million applications, I estimate that they increase the share of applications processed within the time limit by 6 percentage points (11%) and decreased processing times by 13%. The effect is present throughout the 16 months of the experiment and is driven by improvements among offices that were underperforming relative to the median at baseline.

Despite their effect on processing times, the scorecards did not decrease bribe payments. I collect survey data on bribe payments from applicants; the point estimate for the effect on my preferred measure of bribes paid is an increase of BDT 940 (USD 11 or 15%). The lower bound of the 95% confidence interval is a decrease of 4%, thus the result is inconsistent with models where faster service delivery reduces corruption. Using a randomized information intervention among surveyed applicants, I rule out that lack of information about the improved processing times is the reason for not seeing a negative effect on bribes paid.

The positive effect of the scorecards on bribe payments is concentrated among the offices that were overperforming relative to the median at baseline, where the scorecards had no effect on processing times. In the underperforming offices, where scorecards improved processing times, there is no effect on bribes. This is inconsistent with any causal relationship between average processing times and bribes, because processing times can improve without bribes changing, and bribes can increase without processing times changing.

I propose a model whose predictions are consistent with the experimental results. In this model, bureaucrats trade off reputation, bribe money, and the utility cost of effort. Their reputation is determined by their visible job performance along two dimensions, processing times and bribes taken. The scorecards are modeled as an increase in the visibility of processing times, making processing times more important for reputation, which in turn incents all bureaucrats to improve their processing times (akin to a substitution effect).[1] However, for overperforming bureaucrats, the increased visibility of their good performance also increases their reputation level, which reduces the marginal importance of reputation (akin to an income effect). For overperforming bureaucrats, the two effects run in opposite directions, and the overall effect predicted by the model on processing times is ambiguous. For underperforming bureaucrats, the two effects run in the same direction, and the model predicts improved processing times.

In the model, bureaucrats refrain from taking more bribes because bribes negatively affect their reputation. When the scorecards improve the reputation level of overperforming bureaucrats they reduce the marginal importance of reputation for their utility and cause them to take more bribes. For underperforming bureaucrats, the decrease in reputation from their performance being more visible is counteracted by their increased effort, so the overall effect of the scorecards on reputation, and therefore also on bribes, is ambiguous.

This paper contributes to four strands of literature. First, it provides empirical evidence on the causal relationship between the average speed of public-service delivery and corruption. It has been shown that slow public-service delivery is positively associated with corruption (Kaufmann and Wei, 1999; Freund et al., 2016), and that individuals seeking services may have to pay bribes to reduce the time between application and provision (Bertrand et al., 2007). In the theoretical literature, one view is that corruption allows individuals to circumvent excessive bureaucratic hurdles (Leff, 1964; Huntington, 1968). An opposing view is that corruption causes delays and red tape in public services, because making the *de jure* regulation more onerous allows government officials to extract more bribes (Rose-Ackerman, 1978; Kaufmann and Wei, 1999).[2] According to both views, we could reduce corruption by improving the speed of service delivery for everyone. How-

[1] The increase in visibility can be interpreted as an improvement in the supervisors' ability to monitor this aspect of the bureacrats' work, an increase in the saliance of the information to the bureaucrats themselves, or both.

[2] In Banerjee (1997), Guriev (2004), and Banerjee et al. (2012), both corruption and red tape emerge from the nature of public-service provision due to the principal-agent problem between the government and its bureaucrats. The experimental results can neither reject nor provide evidence in favor of these models.

ever, I show that increasing the average speed of service delivery does not decrease bribe payments in this context. Furthermore, I provide evidence against any causal relationship between the average speed of service delivery and bribes.

Second, this paper contributes to the literature on how incentives shape bureaucratic performance and corruption. There is an extensive literature on both monetary and non-monetary explicit incentives (e.g., Duflo et al., 2012; Ashraf et al., 2014; Khan et al., 2016, 2019), and a growing literature on the effects of information systems within government bureaucracies (Dhaliwal and Hanna, 2017; Muralidharan et al., 2020; Callen et al., 2020; Raffler, 2020; Dal Bó et al., 2021; Banerjee et al., 2021; Dodge et al., 2021; de Janvry et al., 2021). Consistent with this literature, I show that increased transparency 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 bureaucrats' career concerns (Niehaus and Sukhtankar, 2013a; Bertrand et al., 2020), or to their own sense of shame or pride (Dustan et al., 2018). Complementing the empirical results, my model suggests that reputational concerns among bureaucrats create incentives for performance and limit corruption.

Third, this paper contributes to the literature on the effects of performance monitoring. The finding that providing performance feedback leads to improvements in performance, especially for underperformers, is consistent with the results of several other experiments (e.g. Allcott, 2011; Ayres et al., 2013; Byrne et al., 2018; Barrera-Osorio et al., 2020), including recent work with Indian bureaucrats by Dodge et al. (2021).[3] Most papers in this literature show smaller or even negative effects for high-performers. This paper adds to this literature by showing that these negative effects can spill over into domains not covered by the performance information, for which data is not typically collected. There is a substantial theoretical and empirical literature on the multitasking problem, i.e., how incentives

[3] Ashraf et al. (2014) and Ashraf (2019) show that privately provided social comparisons reduced the performance of low-performing healthcare and garment workers, while publicly announcing good performances increases performance for low-performing groups. Blader et al. (2020) show similar effects among truck drivers, but in their experiment the results are reversed when the intervention is combined with a management practice focusing on establishing a cooperation-based value system.

for improving one indicator have negative spillovers by taking attention and resources away from other types of performance (Holmström and Milgrom, 1991; Finan et al., 2017). My model suggests a different mechanism for negative spillovers, namely the decreasing marginal utility from reputation after receiving positive performance feedback.[4]

Fourth, this paper contributes to the literature on the determinants of bribe amounts. Some models and empirical evidence suggest that bribe payers' outside options and abilities to pay constrain bribe amounts (Shleifer and Vishny, 1993; Svensson, 2003; Bai et al., 2019), potentially leaving little room for applicant complaints or government monitoring to reduce corruption (Niehaus and Sukhtankar, 2013b). In other settings, such monitoring has been effective in reducing corruption (Reinikka and Svensson, 2005; Olken, 2007). I show that, in my context, individual bureaucrats *can* increase bribes and that bribes are not fully determined by the applicants' willingness to pay. My model highlights how bureaucrats' concern for their reputation constrains bribes, thus explaining why bribes are substantially below applicants' willingness to pay for the service.

The rest of this paper is organized as follows. Section 2 describes the context, experimental interventions, and data. Section 3 describes my empirical strategy. Section 4 presents the effects of the scorecards on processing times and bribes. Section 5 discusses mechanisms and how the results relate to existing theories of the relationship between the speed of service delivery and corruption. Section 6 describes the model of bureaucratic behavior and how it can explain the empirical results. Section 7 concludes.

2 Context, Experimental Intervention, and Data

2.1 Land Record Changes in Bangladesh

This paper studies land record changes (called *mutations* in Bangladesh) and the time it takes to process applications for them. When a parcel of land changes owners, either through sale or inheritance, the official record of ownership has to be updated and a new

[4] This is also consistent with the literature on moral licensing, showing that when past prosocial behavior is made more salient, individuals tend to act less altruistically (Sachdeva et al., 2009; Clot et al., 2018).

record of rights (*khatian*) issued to the new owner. An updated record of land ownership is crucial for maintaining secure property rights. Unfortunately, the burdensome and costly process of applying for land records is causing many new land owners to wait to apply until they need a record of rights. This means that the government land records substantially lag actual ownership, contributing to land disputes, one of the most severe legal problems in Bangladesh—where 29% of adults have faced a land dispute in the past four years (Hague Institute for Innovation of Law, 2018).

2.1.1 Structure of Bureaucracy

Applications for land record changes are processed by civil servants holding the position of Assistant Commissioner Land (ACL). Throughout the paper, I refer to ACLs as the *bureaucrats*. ACL is a junior position in the Bangladesh Administrative Service, the elite cadre of the Bangladesh Civil Service. Each subdistrict (*Upazila*) land office is headed by a single ACL and processing land record changes is a central duty of the ACL. In qualitative interviews, ACLs estimate spending about a third to half of their working time on land record changes. Bureaucrats typically hold the position of ACL for one to two years; when an ACL is transferred, it is typically to a non-ACL position.[5]

The ACL is directly supervised by an Upazila Nirbahi Officer (UNO), the most senior civil servant at the subdistrict level. During periods when no ACL is assigned to an office, the UNO is responsible for the ACL's duties. The UNO has substantial power to influence the ACL's future career: through filing an Annual Confidential Report about the ACL's performance with the Ministry of Public Administration. The UNO is in turn supervised by a Deputy Commissioner (DC), the most senior bureaucrat at the district level. Throughout the paper, I refer to the UNOs and DCs as the *supervisors*.

[5]Of the 617 ACLs I observe in my administrative data, only 10% held the position of ACL in more than one land office in the 60% of land offices covered by the experiment.

2.1.2 Application Process

The *de jure* process for making a land record change is depicted in Appendix Figure A1. The process starts when the new owner applies at the relevant subdistrict land office. There is no competition between land offices for applicants, because each parcel of land is under the jurisdiction of a single subdistrict. The application is 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, which is the lowest tier of land offices. There, a Land Office Assistant verifies the applicant's claim to the land by meeting with the applicant and visually inspecting the land. The Land Office Assistant then sends a recommendation back to the subdistrict land office on whether to accept or reject the application. The application is then verified against the government land record. Finally, the ACL holds a meeting with the applicant where the application is formally approved. The applicant then pays the official fee of BDT 1,150 (USD 14) and receives the new record of rights.[6] The subdistrict land office also changes the official government land record to reflect the new ownership.

The government has mandated that applications should take no more than 45 working days to process, but in practice delays beyond this time limit are common.[7] In my data, only 56% of applications in the control group were processed within the time limit, and the average processing time was 64 working days.

2.1.3 Bureaucrats' Discretionary Powers and Corruption

In practice, applicants also pay bribes. Corruption is widespread in land administration—Transparency International Bangladesh (2018) estimates that the land sector is the second-largest receiver of bribes from Bangladeshi citizens. Appendix Figure A2 shows that among the applicants in my survey, the average estimated bribe payment for a "normal person" like themself is BDT 6,718 (USD 80).

[6] Throughout the paper, I use a USD/BDT exchange rate of 84.3, the average exchange rate during the experiment. This exchange rate is not adjusted for purchasing power parity.

[7] At the end of the experiment, the limit was shortened to 28 working days, but the design of the scorecards never changed. They were discontinued in March, 2020.

Appendix Figure A3 shows that the most common responses to the open-ended question of why a bribe was paid are akin to: "to get the work done" (39%), "to avoid hassle" (38%), and "for faster processing" (9%). This highlights the bureaucrats' power over applications along two dimensions. First, they can decide whether to accept or reject the application. Second, they can speed up or slow down the application, as well as create various hassles for the applicant.

Appendix Figure A2 shows that the average stated valuation of getting the record of rights is BDT 1,610,998 (USD 19,110). This is more than two orders of magnitude larger than the estimate of the average bribe payment. Applicants' stated willingness to pay for having their application processed within the shortest realistic processing time (seven days) is, on average, BDT 2,189 (USD 26). Because this amount is substantially lower than my preferred estimate of the average bribe, it suggests that applicants are paying bribes not just for faster processing, but also for getting the approval.

2.1.4 E-governance System for Land Record Changes

In February 2017, a new e-governance system for land record changes was gradually introduced to simplify the process for both applicants and bureaucrats.[8] The e-governance system generates administrative data on each application made in the system. However, until the start of the experiment, this data was not used for measuring adherence to the 45-working-day time limit or for evaluating specific subdistrict land offices or ACLs.

2.2 Experimental Intervention: Performance Scorecards

The experimental intervention consist of monthly scorecards designed to decrease delays in the processing of applications for land record changes. The content of the scorecards is

[8] The system was gradually implemented in subdistrict and local land offices. Because the system was new at the time of the experiment, even in the subdistrict offices where it had been installed, not all applications were processed through it. The most common reson was because some of the *local* land offices within these subdistricts had not yet installed it. Other reasons cited for using the paper-based system were problems with internet connectivity, new officials not yet trained to use the system, and temporary problems with the e-governance server.

automatically generated using data from the e-governance system. While the intervention is randomized at the land-office level, the scorecards were addressed to the bureaucrats (ACLs) and sent to each bureaucrat's land office, as well as to the offices of the supervisors (the UNO and DC). Two versions of the scorecard, one in English and one in Bengali, were sent out by courier in the first two weeks of each month.[9] 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 phone calls to the ACLs, where it was confirmed that the scorecard had been received and where the indicators were explained.

Appendix Figure A5 depicts an example scorecard. The scorecard evaluates the bureaucrat's performance using two performance indicators: the number of applications processed within 45 working days in the past month, where a higher number indicates a better performance, and the number of applications pending beyond 45 working days at the end of the month, where a lower number indicates a better performance.[10] The scorecards compare these numbers with the average numbers for all land offices in the experiment.[11] The scorecard also provides the office's percentile ranking for each indicator, with a short sentence and a thumbs-up or thumbs-down symbol reflecting the performance. Appendix B.3.1 discusses the rationale for choosing these performance indicators and analyzes the extent to which bureaucrats matter for the performance score.

2.3 Randomization and Implementation Timeline

Figure 1 depicts the randomized interventions and data collection. The scorecard intervention was carried out in two waves; each wave randomized all offices where the e-governance system had been installed at that time into either the treatment group or the

[9]An email with a PDF version of the scorecard was also sent to bureaucrats who had listed their email addresses publicly or in the e-governance system. No scorecards were sent in January 2019 due to some bureaucrats' responsibilities the previous month were shifted to the December 2018 elections instead of their regular duties.

[10]The scorecards contained an explanatory note showing how the numbers in the scorecard are calculated and a phone number to call to ask questions.

[11]The initial comparison group was the 112 offices of the first randomization. After the second randomization, the group was expanded to include all 311 offices.

control group. I conducted the first randomization in August 2018: 56 of the 112 land offices where the e-governance system had been installed were randomly selected to receive the scorecards. Appendix B.1 provides details about the stratification of the randomization. The first wave of scorecards started in September 2018. By April 2019, 199 additional offices had installed the e-governance system. 99 of these offices were randomly selected to receive the performance scorecards. The second wave of scorecards started in April 2019. The scorecards were sent out monthly until March 2020, when the COVID-19 pandemic caused the intervention to end.

2.3.1 Additional Intervention: Peer Performance List

To test for peer effects, a Peer Performance List was added to the scorecards for 77 randomly selected offices already receiving the scorecards. The addition took place in September 2019, a year after the first scorecards were sent out. The list contains the percentile rankings of the two performance indicators for all 77 offices. The list also told the bureaucrats that the list had been sent to the 76 other offices, making it clear that their peers could observe their performance. Appendix Figure A6 shows an example of such a list.

2.4 Data

I use two main data sources, administrative data from the e-governance system for all land offices in the experiment and data from a survey conducted among applicants in the offices that were part of the first randomization wave. Table 1 shows summary statistics for both data sets. This table contains all observations from both treatment and control offices used in the analysis. For a discussion of the balance of randomization, see Section 2.6.

2.4.1 Administrative Data

The administrative data is based on 1,034,688 applications from all 311 land offices. The data contains information about the office in which the application was made, the application start date, the date it was processed, the decision to accept or reject the application, and the size of the land.[12] The administrative data was downloaded from the e-governance system at the beginning of each month from August 2018 until December 2020.

For the main analysis, I use administrative data for applications made from August 13, 2018, one month before the start of the intervention, until January 20, 2020. I include applications made one month before the intervention if they had not been processed by the arrival of the first scorecard, as these were partially treated. I chose to end the data 45 working days before March 26, 2020—this is when the COVID-19 pandemic caused a long general holiday and the end of the intervention. Choosing this cutoff point precludes the general holiday from affecting whether the applications were processed within the 45-working-day time limit.

The processing time is measured as the number of working days between the application start date and the date the processing was finalized. I impute the processing times for the 2% of applications that had not yet been processed by December 2020.[13] The imputed value is the mean of actual processing times that were longer than the time the application had been pending. Appendix B.3.2 provides more information about the administrative data.

2.4.2 Survey Data

I collected the survey data in two rounds from applicants who applied in the 112 offices that were part of the first randomization wave. To create the sample of applicants, enumerators were stationed outside land offices to interview all applicants entering the office for the purpose of a land record change, regardless of their stage in the application process. The enumerators stayed outside a specific office for at least two days and until they had completed at least 20 interviews. This first-round interview focused on the basic details of the application and applicant, the applicant's expectation for the application processing

[12]The full administrative data set also contains more information about the applicants, such as names and phone numbers, but this data is not available for research purposes due to privacy concerns.

[13]The estimate of the scorecards' effect on the share of not-yet-processed applications is a decrease of 0.4 percentage points.

time, and the applicant's willingness to pay for faster processing.

The follow-up interviews, conducted by phone approximately three months after the initial interview, focused on the outcome of the application and bribe payments. Enumerators were not informed about which offices had received the scorecards or whether they were calling a respondent from a treatment or control office.

To avoid extreme outliers potentially caused by enumeration errors, all continuous variables from the survey are winsorized at the 99th percentile. Appendix B.3.4 provides more information about the survey data.

There are two measures of bribe payments. The first is based on a question of how much the applicant thinks it is "normal for a person like yourself to pay." For the 63% of respondents who were willing to answer this question, the amount is recorded as the variable *typical payment.* The average response is BDT 6,718 (USD 80) or 1.5 months of the sample's average per capita household expenditure. 73% of the responses were nonzero amounts. The second measure is based on a series of questions about each applicant's actual payments to any government officials or agents assisting with the application. The outcome variable *reported payment* is the sum of the reported amounts. The average reported payment is BDT 1,477 (USD 18), and 27% of respondents provided a nonzero value. Among those reporting a nonzero amount, the average amount was BDT 5,283 (USD 63).

The typical payments measure is my preferred measure of bribes. The main problem with the reported payment measure is the large number of zero responses, suggesting that it is an underestimate of bribes paid. However, I have no reason to believe that either of the two measures is biased differently between the treatment and control offices. Throughout the paper I show that the main results are robust to using either of the two measures.

Of the 3,213 applicants from the in-person interviews, 2,869 were successfully interviewed in the follow-up phone call, for a total attrition rate of 11%. The estimated effect of the scorecards on the attrition rate was 3 percentage points (p-value = 0.08). In Appendix B.3.6, I discuss attrition in detail and use Lee bounds (Lee, 2009) to show that the differential attrition is not sufficiently large to substantially affect the main findings from the survey data.

2.5 Additional Intervention: Providing Information to Applicants

An additional experimental intervention giving applicants information about processing times was also carried out together with the survey. The motivation for this intervention was to ensure that applicants knew about the improvements in processing times. Though this information would likely have spread eventually, in the short-term, information about changes to bureaucrat behavior may not yet have been disseminated. To speed up this dissemination and approach the long-term effect of the scorecards faster, on randomly selected days the enumerators gave applicants leaflets that described how 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 implemented. The information was the same in treatment and control offices.

Appendix Figure A7 shows an English translation of the leaflet. Due to some non-compliance with the treatment assignment by the enumerators delivering the treatment, I use the median treatment delivered in a land office survey day as the main treatment variable when analyzing the effect of the information intervention. Appendix C.4.1 discusses this choice and shows the robustness of the main results to using alternative treatment variables.

2.6 Balance of Randomization

Appendix Table A1 shows balance-of-randomization tests for variables from the administrative data. To exclude all data that the scorecards could have affected, I restrict the data to applications made at least 45 working days before the start of the experiment. Applications not processed by the start of the experiment were assigned an imputed processing time based on the time they had been pending at the start of the experiment, using the imputation procedure described in Section 2.4.1. There are no statistically significant differences between scorecard and control offices before the start of the experiment. This is expected, given the random treatment assignment.[14]

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 find no evidence for any such changes in composition. Comparing the age and income of the applicants, the size and value of the land the applications are for, and the stages that the applications are in at the time of the first interview there are no statistically significant differences.[15]

3 Empirical Strategy

3.1 Empirical Strategy: Overall Effects

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

$$Outcome_{ait} = \alpha + \beta Treatment_i + Stratum_i + Month_t + \varepsilon_{ait} \tag{1}$$

where $Outcome_{ait}$ is an outcome for application a, in land office i, made in calendar month t. $Stratum_i$ are randomization stratum fixed effects. Since no randomization stratum overlap the two randomization waves, these fixed effects also control for randomization-wave fixed effects. $Month_t$ denotes fixed effects for the month the application was made.[16] For the main results, I provide p-values testing the null hypothesis of no effect using conventional standard errors clustered at the land-office level, as well as p-values based on

[14]Making the same comparisons seperately for offices over- and underperforming at the start of the experiment similarly yields no statistically significant differences.

[15]Making the same comparisons for offices over- and underperforming at the start of the experiment yields one difference statistically significant at the 5% level: among overperforming offices, a smaller share of applications were approved at the time of the in-person interview in the treatment offices. One more difference is statistically significant at the 10% level: among overperforming offices, a smaller share of applicants are female in the treatment offices. Overall, I perform 27 balance tests in the survey data of which only one is significant at the 5% level, and an additional one at the 10% level.

[16]In the survey data, consistent with the cleaning of other continuous variables, the application month variable is winsorized at November 2018, so that all application dates in or before November 2018 take the same value. A separate indicator variable controls for missing start-date values.

randomization inference.[17] Each observation is weighted by the inverse of the number of observations in each land office. This has three benefits. First, it makes the regression estimate the average effect of the scorecards on a land office, the unit relevant for studying changes in bureaucrat behavior. Second, using these weights, the analysis of the administrative data and the survey data estimates the same effect. Third, the weighting improves the estimates' precision by weighting each cluster equally in the analysis.[18]

The two additional randomized interventions—the peer performance lists and the information intervention to applicants—are excluded from the main specification, because they are not the main treatments being evaluated. For the two main outcomes—processing times and bribe payments—the full specifications, including the scorecard treatment, the relevant additional randomization, and the interaction, can be found in Appendix Tables A3 and A4. These tables show that neither of the two additional experiments have substantial interactions with the scorecard treatments, validating my approach to analyze the scorecard treatment separately.

3.2 Empirical Strategy: Heterogeneous Effects by Baseline Office Performance

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.[19] I define baseline performance as the average of the two performance indicators' percentile rankings in the first month of treatment. I then separate all offices into *overperformers* (above the median baseline perfor-

[17] The randomization inference is implemented using the Stata command *randcmd*, and the reported p-value is from the randomization t-test calculated using 9,999 iterations (Young, 2019).

[18] 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.

[19] Heterogeneity in the effects of performance-information provision between high and low performers has been recorded in several settings (e.g., Allcott, 2011; Ashraf, 2019; Barrera-Osorio et al., 2020; Dodge et al., 2021). This was the only heterogeneity test based on office characteristics specified in the preanalysis plan. The two other prespecified 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 appear in Appendix Figure A8 and Appendix Table A5, respectively.

mance) and *underperformers* (below the median baseline performance).[20] Since the classification of offices uses data only 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 overperforming at baseline, β_2 is the effect for offices underperforming at baseline, and γ is the difference between overperforming and underperforming offices in the control group.[21] As in the estimation of the overall effects, standard errors are clustered at the land-office level and the regressions are weighted by the inverse of the number of observations in each land office.

4 Results: Effects of the Scorecards on Processing Times and Bribes

4.1 Effect on Processing Times

Table 2 shows that the scorecards increased the number of 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 whether the application

[20] I classify 112 offices in the first randomization wave into over- and underperformers by comparing them to the median performance among these 112 offices. 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. This makes the *overperformer* and *underperformer* classifications correspond to the relative performance presented in the first scorecards.

[21] To test the hypothesis that the treatment had the same effect on offices overperforming and underperforming at baseline, I use a similar regression but where the first treatment variable is not interacted with the indicator variable for whether the office overperformed at baseline. I then test the hypothesis that the coefficient on the treatment variable interacted with with the indicator variable for whether the office was underperforming at baseline is zero. This test's p-value is reported as "P-value subgroup diff." in the regression tables reporting the heterogeneous effects.

was processed within the 45-working-day time limit. The scorecards increased the share of applications processed within the limit by 6 percentage points (11%). Column (2) shows that the scorecards led to a 13% reduction in overall processing times by estimating the effect on the natural logarithm of the processing time.[22]

For column (3), I create a *time index* of the two outcomes used in columns (1) and (2).[23] The estimated effect of the scorecards on the time index is 0.13 standard deviations. Using conventional clustered standard errors, the rejection of the null hypothesis has a p-value of 0.028; using randomization inference, the p-value is 0.037.

4.1.1 Effect on the Distribution of Processing Times

Figure 2 shows the effect using minimally processed data. It 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 includes only applications that have already been processed, and processing times are top-coded at 180 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 no substantial bunching just before the 45-day limit. This is to be expected given that the process for approving an application is relatively long and depends on several individuals, as described in Section 2.1. Thus, even if the bureaucrat cares only about maximizing the share of applications processed within 45 working days, the processing time target has to be 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 emphasize both processing applications within the 45-working-

[22] The exact effect is -12.6 log points, which is equivalent to an 11.9% decrease. For simplicity, I will describe log point changes as percentage changes throughout the paper.

[23] I created the index by first taking the negative of the log processing time so that a higher value indicates a better performance, then recasting the two outcome variables as standard deviations away from the control-group mean, and finally taking the sum of the two standard deviations and rescaling them so that the index has a standard deviation of one in the control group.

day limit and reducing the number of applications pending beyond the limit. Overall the spread of the effect in the distribution of processing times alleviates the concern that bureaucrats are "gaming" the scorecards by speeding up the processing only of applications that would otherwise have been processed just outside the time limit.

4.1.2 Effect Over Time

It is possible that the initial effect of the scorecards is different from the long-run effect. If there was a substantial novelty effect, we would expect to see a decline in the difference between the treatment and control groups over the 16 months of the intervention. Figure 3 and Appendix Figure A8 show no pattern of the effect declining over time, though the size of the effect varies between different periods.

Figure 3 shows the 10-working-day moving average of the share of applications processed within the time limit by application date relative to the start of the experiment. The first and second vertical lines indicates the date 45 working days before the first scorecards were sent and the date of sending them, respectively. Applications made between these dates were partially affected by the scorecards, as the bureaucrats received the scorecards while processing these applications. Starting just before the first scorecards were sent, we see that the treatment group process more applications on time relative to the control group. With a few short exceptions, the treatment offices continued to process a higher share of applications within the time limit until the end of the experiment. Relative to the first scorecards, the data from the offices in the second randomization wave ends earlier. The third vertical line marks where the graph contains data only from the offices in the first randomization wave.

Appendix Figure A8 shows the results of applying the regression specification from Equation 1 to applications made in the first, second, and final third of the experimental period. The outcome variable is the time index from column (3) of Table 2. There is no pattern of a continuous decline of the effect over time.[24]

4.2 Effect on Bribe Payments

Table 3 shows that the scorecards did not lead to a decrease in bribe payments. Instead, the estimated effect on bribes is positive, though this increase is not statistically significant. As described in Section 2.4.2, data on bribe payments was collected using two separate survey questions. The first question asked about the typical bribe payment "for a person like yourself." When this measure is used, the column is marked as *typical*. The second question asked about each payment made by the applicant. When this measure is used, the column is marked *reported*.

Columns (1) and (2) of Table 3 show the effect on the amount of bribes paid. Column (1) shows that the effect on the perceived typical payment was BDT 940 (USD 11), a 15% increase. The rejection of the null hypothesis has a p-value of 0.130 (randomization inference p-value = 0.159). While the 95% confidence interval contains a zero effect, the lower bound of the confidence interval is a decrease of 4%, ruling out a meaningful decrease. Column (2) estimates that the scorecards increased reported bribe payments by BDT 297, a 23% increase (conventional p-value = 0.105). Columns (3) and (4) show that there is no effect on the propensity to report a nonzero 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 nonzero bribe payments. Bribe payments increased by 18% for typical payments (conventional p-value = 0.053) and 25% for reported payments (conventional p-value = 0.011).[25]

[24] Estimating the effect on the time index for applications made after the end of the intervention (March 26, 2020) until 45 working days before the the end of my data (September 28, 2020) yields an estimate of 0.07 standard deviations. This suggests that there is only a small amount of persistence in the effect of the scorecards. However, this estimate should be interpreted carefully—the COVID-19 pandemic drastically changed the operating conditions for the bureaucrats during this period.

[25] Again these effects have two interpretations. Either the scorecards affected only 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.

4.3 Heterogeneity of Results by Baseline Office Performance

Table 4 uses the empirical strategy from Section 3.2 to show that the scorecards' effect on processing times is driven by offices that were underperforming at baseline, while the effect on bribe payments is driven by offices that were overperforming at baseline. Column (1) of Table 4 shows that for offices that were underperforming at baseline, the estimated effect on the time index is an increase of 0.24 standard deviations. Using conventional clustered standard errors, the rejection of the null hypothesis has a p-value of 0.007 (randomization inference p-value = 0.012).[26] For offices that were overperforming at baseline, the effect is just 0.03 standard deviations.

Columns (2) and (3) in Table 4 show that the positive effect on bribe payments is entirely driven by the offices that were overperforming at the start of the experiment. Column (2) shows that the effect of the scorecard on estimated typical bribe payments among offices overperforming at baseline is an increase of BDT 2,069 (38%). Using conventional clustered standard errors, the rejection of the null hypothesis has a p-value of 0.008 (randomization inference p-value = 0.016).[27] The effect on offices that were underperforming at baseline is close to zero.

Appendix Tables A7 and A8 show that the heterogeneous effects on processing times and bribes by baseline performance remains similar after controlling for other office baseline characteristics interacted with the treatment. However, the association between the effect size and baseline performance should still not be interpreted as a causal relationship. Baseline performance is not randomly assigned, and it is plausible that unobserved office characteristics are associated with both performance and the treatment effect size. Instead, I interpret the results as describing which type of offices reacted in what ways to the scorecards.

Even without a causal interpretation, the heterogeneity in the results is surprising:

[26] The randomization p-value is 0.023 when using the Westfall-Young multiple-hypothesis testing method to adjust for that I am testing two hypotheses (Westfall and Young, 1993; Young, 2019).

[27] The randomization inference p-value is 0.032 when using the Westfall-Young multiple-hypothesis testing method to adjust for that I am testing two hypotheses.

overperforming offices did not change their behavior in terms of processing times, but among these offices bribe payments increased. Sections 5 and 6 investigate this further.

4.4 Effects on Visits, Processing Times, and Satisfaction in Survey Data

In Appendix Table A6, I use survey data from the 112 offices in the first randomization wave to show that the scorecards reduced the number of visits to land offices, increased the share of applications processed within 45 working days, and decreased overall processing times.[28] The results on processing times are not statistically significant but the heterogeneity between offices over- and underperforming at baseline are consistent with the results from the administrative data. Appendix C.6 analyzes survey data on applicants' stated satisfaction with the application process and shows that the scorecards did not improve the applicants' stated satisfaction. A negative, but not statistically significant, effect is found for applicants in offices overperforming at baseline.

4.5 Robustness Tests

The main results for the effects of the scorecards on processing times and bribes, as well as the heterogeneity in these effects, is robust to a range of alternative specifications. Appendix Tables A9 and A10 show the results using various combinations of controls, weights, and winsorizations of the bribe amounts, as well as including only bribes given directly to government officials while ignoring fees paid to agents. All alternatives to the main estimates are of the same sign and of similar magnitude, but some of them are not statistically significant. Appendix Table A11 shows the effects, estimated at the office-month level, on the number of applications processed within 45 working days and 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, driven by improvements in offices underperforming at baseline,

[28] Based on these three outcome variables, I also construct an inverse-covariance weighted (ICW) index using the algorithm suggested by Anderson (2008). The index is designed to summarize several outcome variables into one variable while weighing the variables in a way that maximizes the information captured.

but only the results for underperforming offices are statistically significant. Appendix Table A12 shows that the estimated effect on processing times is robust to different functional form assumptions and different imputation techniques for applications that were not yet processed.

Appendix Table A13 shows that the heterogeneity found in the effects is robust to three alternative measures of baseline performance. First, I use three months of baseline data to establish the baseline performance—the results are similar to when using one month of baseline data. Second, I separate offices by the performance quartiles and create four, instead of two, performance groups. The positive effect on the processing time index is monotonically decreasing in being in a better baseline performance group. The effect on bribes is monotonically increasing in being in a better performance group, except for in the offices between the 50th and 75th baseline percentile ranking where the estimated effect is slightly larger than in the group of offices above the 75th baseline percentile ranking. Finally, I show that the continuous baseline performance ranking is negatively associated with the size of the effect on the time index but positively associated with the effect on bribe payments, though the latter association is not statistically significant.

It is possible that the survey and information intervention changed the overall effect of the scorecards—for example, if bureaucrats pay attention to the scorecards only if they know that some of their applicants are being surveyed. In Appendix C.1, I describe the results when restricting the sample to applications made before the survey took place and applications made in offices where there was no survey. The estimated effect using these applications is similar to the main estimate, showing that the survey did not substantially alter the treatment effect.

4.6 Unintended Consequences

A common problem of quantitative performance measures is that they often lead to gaming of the measures or other unintended consequences. In Appendix C.2, I test for four such potential unintended consequences. One, if bureaucrats allow fewer applicants to start applications, then their scorecards may improve, provided that the lower number of applications helps them process a larger share of the applications within the time limit. Two, if bureaucrats allowed applications selectively, such that the average application was easier to process within the time limit, then their scorecards may improve. Three, the scorecards may lead bureaucrats to make worse decisions regarding accepting or rejecting applications. Four, bureaucrats may divert attention from applications not made in the e-governance system, because those applications do not count toward the scorecards. I find no evidence for large unintended consequences, except for suggestive evidence of a higher incorrect rejection rate in offices overperforming at baseline, potentially as a response to applicants not being willing to pay the new higher bribes.

5 Mechanisms and Implications for Existing Theories

5.1 Two Potential Mechanisms for the Effect on Processing Times

The scorecards could improve performance of bureaucrats through two main channels. First, the information that supervisors get may improve their ability to incent bureaucrats by facilitating better promotions and more attractive postings for those bureaucrats with good scorecards. This is an example of the widely studied mechanism of increased information enabling better contracts that improve output (Holmström, 1979).[29]

Second, bureaucrats may change their behavior due to receiving the scorecards themselves. For bureaucrats, receiving information about their processing times each month may increase this information's salience, causing it to be more important for their personal sense of pride in their work. 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.*

[29] 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.

5.2 Effect of Peer Performance List

To investigate another potential mechanism, peer effects, I use a randomly implemented variation of the scorecard that informs bureaucrats at the same level within the bureaucracy about each other's performance (Mas and Moretti, 2009; Blader et al., 2020). It is possible that information flows about performance *between* bureaucrats at the same level in the organizational hierarchy create an additional incentive for improved performance. I measure such a peer effect, as an addition to the effect of the scorecard, by estimating the effect of the peer performance list sent to a randomly selected subgroup of the offices receiving scorecards, as described in Section 2.3.1.

Appendix Table A4 shows no substantial effect of the peer performance list on processing times. Using only data from land offices receiving performance scorecards, columns (1) and (3) shows that the effect on processing times is close to zero. Columns (2) and (4) use the full data set and estimate the effect of being in the the scorecard and the peer performance list treatment groups simultaneously, both before and after the performance list intervention started. This is done using an indicator variable for the peer performance list treatment that takes the value of one for applications made in offices receiving the peer performance lists and a *Post* indicator variable for applications made later than one month before the peer performance list intervention started. The estimated effect is a minor improvement in performance that is not statistically significant.[30]

5.3 Implications for the Relationship Between Processing Times and Bribes

There are several reasons why bribes may be causally related to processing times. In the literature on this relationship, two opposing views lead to drastically different policy conclusions. One view is that bribes "grease the wheels" by providing incentives to bureaucrats and allowing for excessively onerous red tape to be circumvented (Leff, 1964; Huntington, 1968). In this view, rooting out corruption would decrease the speed of service delivery

[30] The effects from the standard scorecard treatment as measured by the coefficients on *Scorecard* and *Post* × *Scorecard* can account for virtually all of the effect estimated by my main specification in Table 2, showing that the effect of the scorecards is not driven by the inclusion of the Peer Performance Lists.

and increase inefficiencies of excessive bureaucratic control. On the other hand, improving the speed of service delivery through better technology or more personnel would decrease bribery, because the need for bribes would decrease.

Another view is that opportunities for corruption is the original reason for excessive red tape and delays in public services—making more-onerous hurdles allows government officials to extract more bribes in exchange for avoiding these hurdles (Rose-Ackerman, 1978; Kaufmann and Wei, 1999). Under these circumstances, providing the bureaucracy with more personnel or resources would not improve processing times, because the bureaucrats would be intentionally slowing service delivery. However, service delivery could be improved by eliminating corruption. The scorecard experiment was originally designed to test a model of this type. See Appendix A.2 for a more detailed description of the model and an empirical test rejecting it.

A common prediction for both types of theories is that faster average service delivery would reduce bribes. However, the results in Section 4 are not consistent with a causal relationship between average processing times and bribe payments. While the scorecards did reduce processing times, they did not reduce bribes, as shown in Tables 2 and 3, respectively. This is true even for the offices that were underperforming at baseline and improved their processing time the most, as shown in Table 4. Furthermore, as bribes increased in overperforming offices, there was no corresponding change in processing times among these offices.

5.3.1 Does Information About Faster Processing Times Decrease Bribe Payments?

One potential reason for the lack of a negative effect from the faster processing times on bribe payments could be that the information about the improvement in processing times had not yet disseminated among applicants when the survey was carried out. The information treatment was designed to test this hypothesis, by informing applicants about improvements in processing times, is described in Section 2.5. Column (2) of Appendix Table A14 shows that the effect of this intervention on expected processing times at the time of the in-person survey was a reduction of 5%.[31] Appendix C.4 shows that the information treatment did not affect bribes, neither by itself nor in combination with the scorecards. Together these two results show that even when applicants are informed about the improved processing times, faster processing times do not reduce bribe payments.

5.3.2 Implications for Other Theories

The increase in bribes among the offices that were overperforming at baseline is also inconsistent with models where it is an applicants' outside option or ability to pay that determines the bribe levels (Svensson, 2003; Niehaus and Sukhtankar, 2013b). If the bribe level was fully determined by the applicants' outside option or ability to pay, it could not have been increased by a positive scorecard, without any observable change in service quality. However, this result is most likely dependent on the market 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 for this service. Therefore, the bribe level is expected to be determined mainly by other factors. Had there been competition for applicants between land offices, or a close alternative to a land-record change, it is plausible that these outside options would have been more important in determining the bribe level (Svensson, 2003; Bai et al., 2019).

A different class of models exists, in which government officials could extract more bribes if they wanted to, but they choose not to do so because they face a trade-off between taking bribes and achieving some other objective. This trade-off could be between bribe money and the risk of getting caught (Becker and Stigler, 1974; Olken, 2007; Niehaus and Sukhtankar, 2013a), but it could also be between bribes and altruistic motivations or reputation concerns. In the next section, I will develop such a model and show that it is consistent with the results of the experiment.

[31] Appendix C.4.1 discusses the noncompliance with the treatment assignment in the delivery of this intervention and shows the robustness of the results to alternative definitions of the treatment variable.

6 Model of Bureaucrat Behavior

In this section, I outline the model I propose to explain the results of the experiment and provide an intuition for its predictions. Appendix A.1 presents the formal model.

6.1 Model Setup

6.1.1 The Bureaucrats' Objective Function

In the model, bureaucrats get utility from a reputation term, which is a function of bribe money taken and visible job performance in terms of processing times.[32] Bureaucrats get disutility from effort, but effort is needed to process applications within the time limit, which improves the bureaucrat's reputation. Bribe money has decreasing marginal utility, while effort has increasing marginal disutility. There is decreasing marginal utility from reputation in performance and increasing marginal disutility through the reputation mechanism in bribe money taken. The intuition for why bribes reduce the bureaucrat's reputation is that for bureaucrats consistently asking for higher-than-normal bribes, rumors of this will spread, even if the bureaucrat is never officially reprimanded. Similarly, if the bureaucrat consistently outperforms peers in terms of processing times, this will build good reputation over time.

The scorecards are modeled as an increase in the visibility of processing times. The scorecards thereby increase the importance of processing times for the bureaucrats' reputation.

I assume that bribe money and visible performance are complements in generating reputation, or equivalently, that honesty (the absence of bribe-taking) and performance are substitutes. In terms of bureaucrats' career prospects, this assumption is based on the idea that some corruption is acceptable as long as bureaucrats are performing their duties well and that poorly performing bureaucrats still have a good career in the bureaucracy,

[32] The reputation term captures both possible mechanisms for the scorecards' effect on processing times described in Section 5.1. I.e., the reputation term represents reasons for why the bureaucrats care about what their supervisors think of them, as well as psychological reasons that are internal to the bureaucrats.

as long as they are honest and follow the official rules. If bureaucrats are both corrupt and poorly performing this could endanger their careers, while honest, high-performing bureaucrats still can't be promoted much faster than their colleagues, because most promotions are based on seniority. An alternative rationale for this assumption is moral licensing, which suggests that the internal shame a bureaucrat may feel from immoral behavior in one aspect of their work can be alleviated by acting better in another aspect.

Finally, I assume that bureaucrats differ only in how much they value their reputation. This could be because of differences in the valuation of future career prospects or differences in intrinsic motivation to be honest and perform well.[33] These differences is what generate over- and underperforming bureaucrats in the model.[34]

6.1.2 Abstractions From Reality

In the model, the applicants simply pay the bribe amount that the bureaucrats are demanding. While this is clearly a simplification, Appendix Figure A2 shows that even the largest estimate for the average bribe is just 0.1% of the average stated value of the record of rights. This difference between applicants' valuations and the amounts they paid suggests that 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 bureaucrats buying reputation or career advancement using money. This avoids bureaucrats taking as-high-as-possible bribes, then use the money to regain their reputation. The results would be the same if bureaucrats would pay for a position as an ACL in the first place but then could not bribe their way to future

[33] Prendergast (2007) and Hanna and Wang (2017) provide evidence that differences in the intrinsic motivations of government officials can be important for public-service delivery and corruption. Bertrand et al. (2020) provide evidence for the importance of differences of future career prospects.

[34] The predictions of the model still hold even if the differences are driven by other bureaucrat characteristics such as ability. I choose to model the differences as in the valuation of reputation because it is consistent with the observation that overperforming bureaucrats collect fewer bribes than underperforming bureaucrats in the control group, as shown in Table 4.

career advancement.[35]

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. Figure 4 depicts the model's predictions compared to the empirical results. Appendix A.1 provides the formal derivations and statements of the predictions.

6.2.1 Effects of the Scorecards on Processing Times

The first set of predictions is for the effect of the scorecards on processing times. The scorecards have two effects on processing times, an *incentive effect* and a *reputation-level effect*. These effects are akin to the substitution and income effects from a wage increase in a labor-supply model. The incentive effect leads to an improvement in processing times for all bureaucrats. This is because the scorecards increase the visibility of the bureaucrats' performance in terms of processing times and therefore the marginal effect it has on utility. This causes bureaucrats to provide more effort to improve processing times.

For overperforming bureaucrats, the reputation-level effect from the scorecards on processing times is negative because the scorecards increase their reputation by making their good performance more visible. Since reputation has decreasing marginal utility, this decreases the marginal utility of reputation and reduces the optimal amount of effort overperforming bureaucrats exert to process applications on time. As the incentive effect and reputation-level effect move in opposite directions for overperforming bureaucrats, the overall effect of the scorecards on processing times is ambiguous and depends on which of the two effects is stronger.

For underperforming bureaucrats, the reputation-level effect from the scorecards on processing times is positive because the scorecards highlight their poor performance and

[35] Weaver (2021) analyses the effects of such bribes in the allocation of job applicants to positions in public-service delivery.

lower their reputation level. Hence, the incentive effect and reputation-level effect move in the same direction, and the model predicts that the scorecards will improve processing times among underperforming bureaucrats.

Prediction 1: Scorecards improve processing times for underperforming bureaucrats

Figure 4 shows how the model's prediction is consistent with the scorecards improving processing times for offices underperforming at baseline, while the effect is close to zero for offices overperforming at baseline.

6.2.2 Effects of the Scorecards on Bribes

The second prediction relates to the effect of the scorecards on bribes taken from applicants. In the model, the bureaucrats can increase bribes by simply asking applicants for more money. What constrains bureaucrats from extracting more bribes is the negative marginal effect it has on their reputation. When the scorecards improve overperforming bureaucrats' reputation, the marginal effect bribes have on utility through the reputation channel becomes less negative. This leads to an increase in bribes taken by overperforming bureaucrats when they receive the scorecards.

Underperforming bureaucrats, suffer an initial decrease in reputation from the information provision, but since effort increases in response to the scorecards, the overall effect on reputation could be positive or negative. Therefore, the model does not predict the effect of the scorecards on bribes for underperforming bureaucrats.

Prediction 2: Scorecards increase bribes for overperforming bureaucrats

Figure 4 shows how the model's prediction is consistent with the scorecards increasing bribes paid in offices overperforming at baseline, while the effect is close to zero for offices underperforming at baseline.

6.3 Alternative Explanations for the Effects of the Scorecards

Appendix C.3 discusses four potential alternative explanations for the experimental results. First, that the scorecards increased applicants' willingness to pay bribes or increased the opportunity cost of bureaucrats' time. This explanation cannot explain the results, because it is not the offices improving their processing times that increases the amount of bribes paid. Second, that transfers of overperforming bureaucrats caused the increase in bribe payments among offices overperforming at baseline. This is inconsistent with the estimated effects on bureaucrat transfers in Appendix Table A15 being close to zero. Third, that the supervisors' (UNOs') reputation level increased, causing them to demand higher bribes. This is unlikely, because land-record changes constitute a much smaller share of supervisors' responsibility than they do for the bureaucrats (ACLs). Forth, that bureaucrats use the scorecards in negotiations with applicants as "proof" that they are able to process the application quickly. This is not consistent with applicants in overperforming offices' expectations on processing times not improving, shown in in Appendix Table A14. It is also inconsistent with the lack of an effect from the information treatment on bribe payments, shown in Appendix Table A3.

7 Conclusion

I have shown that a management information system – providing information about individual bureaucrats' performance – can improve this performance. This effect is present despite the absence of explicit performance incentives, and it is persistent, at least over a 16-month period. The system, made possible by an underlying e-governance system, is an example of how new technologies are creating opportunities for improved management in the public sector.

To create a rough estimate for the value of the improved processing times, I multiply the applicants' average stated valuation of having their application processed one day faster by the reduction in the total number of processing days due to the scorecards.[36] For the 155 offices receiving the treatment, the approximate value is USD 9.7 million per year.[37] This value should be interpreted carefully—it relies heavily on the value stated by the applicants, and it is plausible that some applicants interpreted the question as including the value of having the application approved. However, the number is more than two orders of magnitude larger than the cost to implement the scorecards, which was approximately USD 40,000 per year.[38] The overall welfare effect of the intervention becomes less clear when taking into account the effect on bribe payments. Multiplying the effect of the scorecards on typical payments with the number of applications in the treatment area results in an estimate of the effect on total bribes paid of 6.6 million per year.[39] Furthermore, the scorecards have a negative but not statistically significant effect on stated satisfaction. Overall, there is no strong evidence that the scorecards had either a positive or a negative effect on average applicant welfare.

More than half of Bangladesh's land offices took part in the experiment, making it plausible that the results are externally valid within Bangladesh (Muralidharan and Niehaus, 2017). However, while I designed the scorecards in collaboration with the government, they were produced and distributed by a nonprofit research organization. Hence, one should be cautious when extrapolating the results from the experiment to a potential scale-up by the government itself (Bold et al., 2018). If the scorecards were to be scaled-up to all bureaucrats, there would also be a larger effect on the benchmark performance than what occurred in the experiment, where only half of the offices received the scorecards. This would shift the whole distribution of performance percentiles down. According to the

[36] 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, which took place before the application was actually processed. Applicants with an expected processing time of 7 days or less are excluded from the calculation.

[37] To estimate the number of processing days saved per year, I multiply the estimates for the total number of processing days per year and the improvement in the processing time for the average application. See Appendix C.5 for details.

[38] This approximation includes the setup costs and the time the author contributed toward implementing the scorecards, valued at USD 200 per day.

[39] If instead I use the effect on the reported payment, the total increase is USD 2.1 million per year.

model, this general-equilibrium effect would induce more effort and smaller bribe payments than the partial experimental roll-out of the scorecards.

The results and the model have several policy implications. First, they show that improving the speed of public service delivery is not an effective tool for reducing corruption in this context. Two important features of my setting are that there are no close substitutes to the public service and that the bureaucrats can control both the service delivery speed and whether the service is delivered at all. Further research is needed to investigate if service delivery speed is more important for bribes when bureaucrats can only control the speed of service delivery or when closer substitutes to the service are available.

Second, the differential effects of the scorecards on underperforming and overperforming offices suggest that it is especially important to improve information about underperforming bureaucrats. They imply hat the type of recognition systems that are common for bureaucrats—where outstanding performances are recognized without addressing inadequate performances—are ineffective. Positive feedback might still have an overall positive effect, because it may motivate underperforming bureaucrats who want to receive better feedback. But due to the reputation-level effect, the positive feedback is less effective than the negative feedback and can, in some cases, even be counter-productive.

Finally, the model points out a general problem when using information systems to create incentives for socially desirable behavior. If the reputations or self-perceptions of some agents are improved, this may have negative spillovers on all other behavior where reputation is a motivating factor. When evaluating such information interventions, it is therefore important to measure effects on all domains of performance, not just on the performance domain that is measured in the information system, or performances where there could be direct spillovers due to multitasking. This is an especially important insight for government bureaucracies, where compressed wage structures, secure employment, and potential counterproductive incentives due to corruption often make reputational concerns and a sense of pride in one's work 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, 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.

Ashraf, N., O. Bandiera, and S. S. Lee (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior & Organization 100*, 44–63.

Ayres, I., S. Raseman, and A. Shih (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *The Journal of Law, Economics, and Organization 29*(5), 992–1022.

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., M. C. Best, A. Q. Khan, and A. Prat (2021). The allocation of authority in organizations: A field experiment with bureaucrats. *The Quarterly Journal of Economics 136*(4), 2195–2242.

Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (2021). Improving police performance in rajasthan, india: Experimental evidence on incentives, managerial autonomy, and training. *American Economic Journal: Economic Policy 13*(1), 36–66.

Banerjee, A., R. Hanna, and S. Mullainathan (2012). Corruption. In R. S. Gibbons and J. Roberts (Eds.), *The Handbook of Organizational Economics*, pp. 1109–1147. Princeton University Press.

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.

Banuri, S. and P. Keefer (2016). Pro-social motivation, effort and the call to public service. *European Economic Review 83*, 139–164.

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.

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.

Byrne, D. P., A. L. Nauze, and L. A. Martin (2018). Tell me something i don't already know: Informedness and the impact of information programs. *Review of Economics and Statistics 100*(3), 510–527.

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.

Clot, S., G. Grolleau, and L. Ibanez (2018). Moral self-licencing and social dilemmas: an experimental analysis from a taking game in madagascar. *Applied Economics 50*(27), 2980–2991.

Dal Bó, E., F. Finan, N. Y. Li, and L. Schechter (2021). Information technology and government decentralization: Experimental evidence from paraguay. *Econometrica 89*(2), 677–701.

de Janvry, A., G. He, E. Sadoulet, S. Wang, and Q. Zhang (2021). Performance evaluation, influence activities, and bureaucratic work behavior: Evidence from china. Working paper.

Dhaliwal, I. and R. Hanna (2017). The devil is in the details: The successes and limitations of bureaucratic reform in India. *Journal of Development Economics 124*, 1–21.

Dodge, E., Y. Neggers, R. Pande, and C. Moore (2021). Updating the state: Information acquisition costs and public benefit delivery. Working paper.

Duflo, E., R. Hanna, and S. P. Ryan (2012). Incentives work: Getting teachers to come to school. *American Economic Review 102*(4), 1241–78.

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.

Guriev, S. (2004). Red tape and corruption. *Journal of development economics 73*(2), 489–504.

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

Hanna, R. and S.-Y. Wang (2017). Dishonesty and selection into public service: Evidence from india. *American Economic Journal: Economic Policy 9*(3), 262–90.

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

Holmström, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org. 7*, 24.

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

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.

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.

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

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.

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.

Olken, B. A. and R. Pande (2012). Corruption in developing countries. *Annu. Rev. Econ. 4*(1), 479–509.

Prendergast, C. (2007). The motivation and bias of bureaucrats. *American Economic Review 97*(1), 180–196.

Raffler, P. (2020). Does political oversight of the bureaucracy increase accountability? Field experimental evidence from a dominant party regime. Working paper.

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.

Sachdeva, S., R. Iliev, and D. L. Medin (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological science 20*(4), 523–528.

Sanchez De la Sierra, R., K. Titeca, A. Amani Lameke, and A. Jolino Malukisa (2020). Corruption in hierarchies. Working paper.

Shleifer, A. and R. W. Vishny (1993). Corruption. *The Quarterly Journal of Economics 108*(3), 599–617.

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.

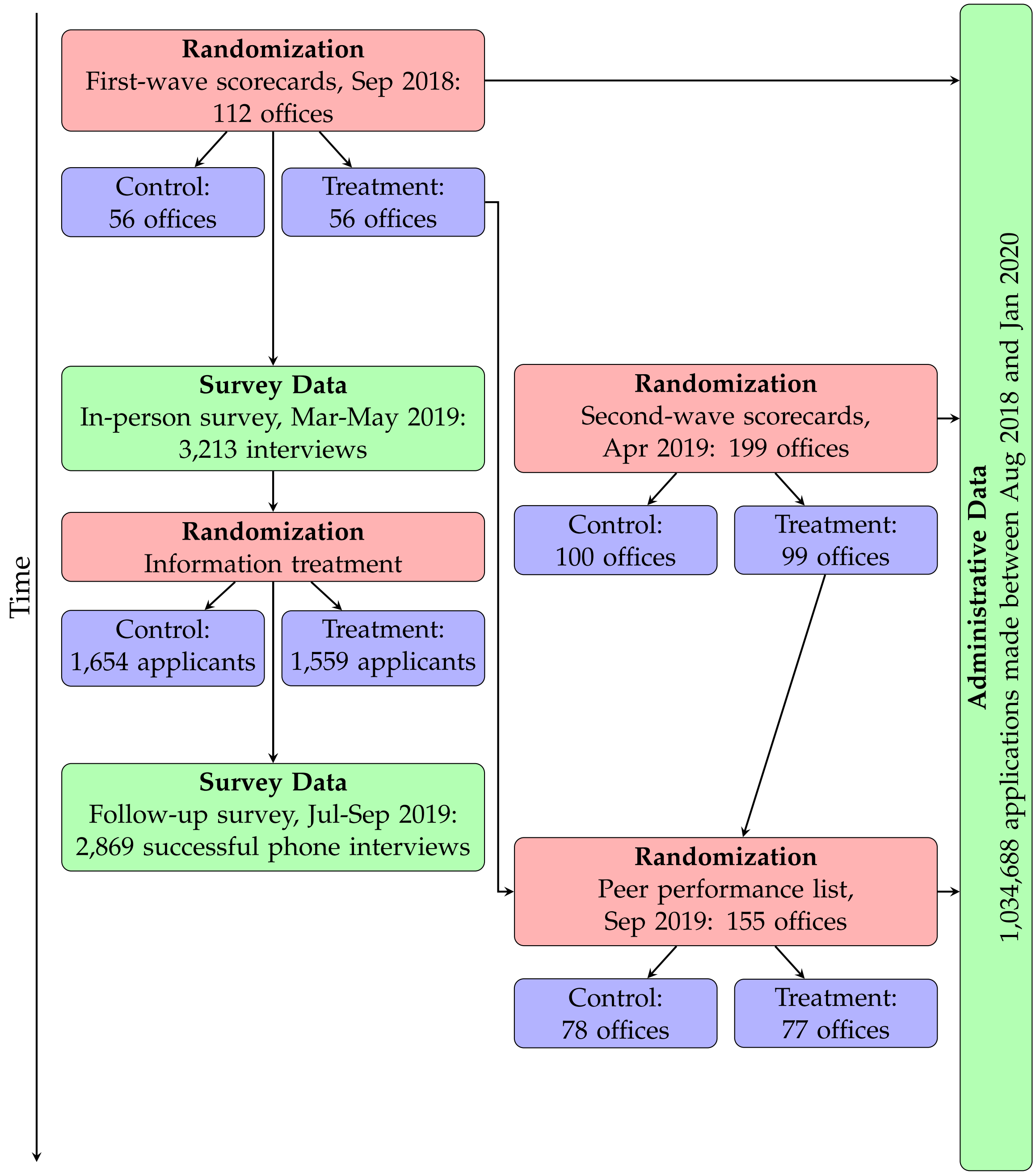
Transparency International Bangladesh (2018). Corruption in service sectors: National household survey 2017. Research report, Transparency International Bangladesh.

Weaver, J. (2021). Jobs for sale: Corruption and misallocation in hiring. *American Economic Review 111*(10), 3093–3122.

Westfall, P. H. and S. S. Young (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*, Volume 279. John Wiley & Sons.

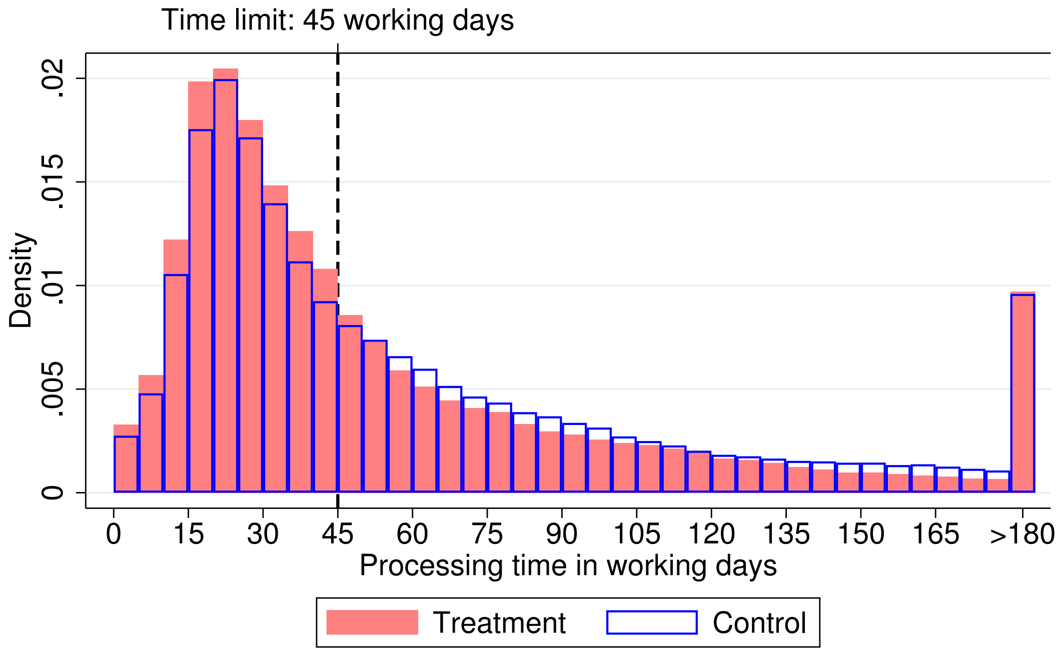
Young, A. (2019). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics 134*(2), 557–598.

Figure 1: Overview of Randomizations and Data Collection



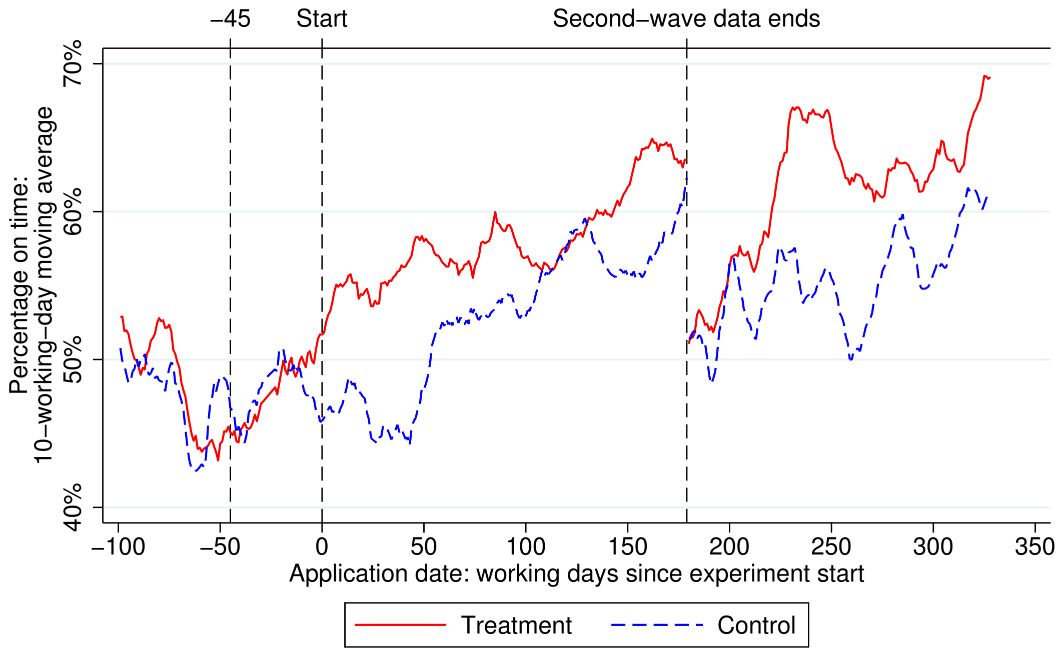
The figure displays the experiment design and data collection. The timeline is chronological from top to bottom. See discussion in Section 2.3.

Figure 2: Histogram of Processing Times by Treatment



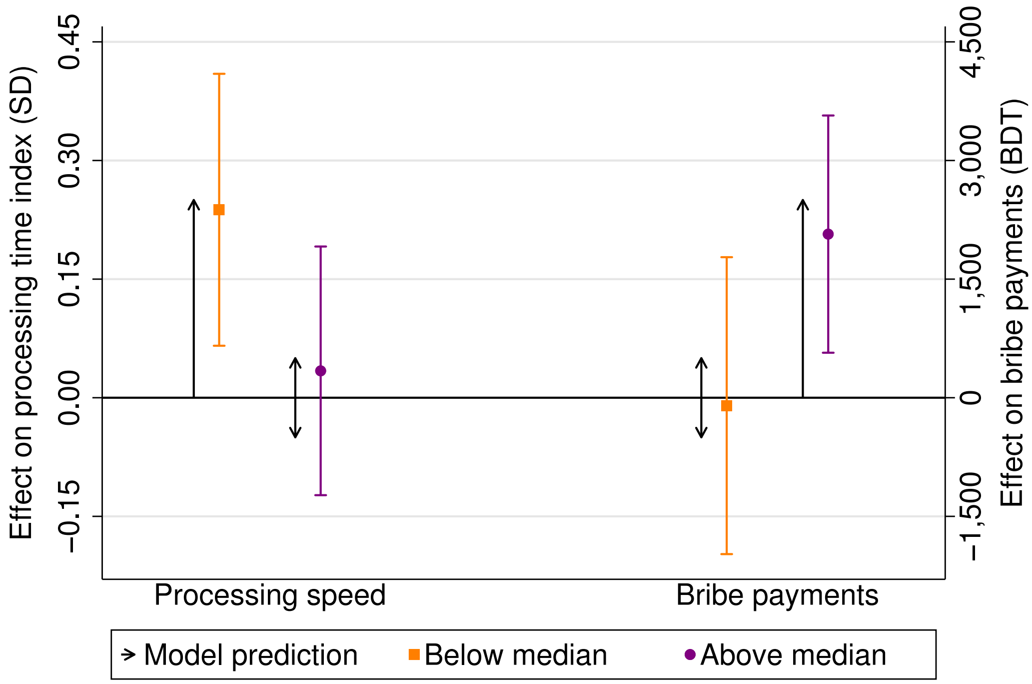
The figure displays histograms of processing times for the treatment and control groups separately. Processing times are top-coded at 180 working days. The data contains 1,006,272 applications made between one month before the start of the experiment and 45 working days before the experiment ended (August 13, 2018 to January 20, 2020). Applications not processed as of December 2020 are excluded. See discussion in Section 4.1.1.

Figure 3: Percentage of Applications Processed Within 45 Working Days



The figure displays the 10-working-day moving average of the percentage of applications processed within the 45-working-day time limit in the treatment and control groups. The data is arranged by the application start dates and contains all applications made from 100 working days before the start of the experiment until 45 working days before the experiment ended. The first and second vertical lines represent the date 45 working days before the first scorecard was sent out and the date of the first scorecard, respectively. Applications made between these dates are partially treated, while applications made after the second line are fully treated. The third vertical line represents the end of the data from the second randomization wave. To the right of this line, the data is only from the 112 offices in the first randomization wave. See discussion in Section 4.1.2.

Figure 4: Model Predictions and Empirical Results



The figure presents the heterogeneous effects of the scorecards from Table 4 compared to the model's predictions from Section 6.2. The arrows represent the direction of the effect predicted by the model for bureaucrats performing above and below the median, respectively. The empirical estimates show the estimated effects for offices performing below and above the median at baseline. The effects on processing speed are measured on the left y-axis in standard deviations of the time index constructed using the variables for whether the application was processed on time and the log of the overall processing time. The effects on bribe payments are measured on the right y-axis in BDT. USD/BDT≈84.3. The bribe data comes from the responses to the question how much it is "normal for a person like yourself to pay". 95% confidence intervals are constructed using standard errors clustered at the land-office level. See discussion in Section 6.2.

Table 1: Summary Statistics

	(1) Mean	(2) Median	(3) St. Dev.	(4) Observations
Panel A: Application-level administrative data				
Process time $\leq$ 45 working days	0.59	1	0.49	1,034,688
Actual process times (working days)	54	35	52	1,006,272
Process time, incl. imputed values (w. days)	61	36	69	1,034,688
Approval rate	0.67	1	0.47	1,006,265
Panel B: Monthly office-level administrative data				
Total applications	282	211	268	4,516
Applications processed	241	136	332	4,516
Apps. disposed within 45 working days	151	79	195	4,516
Apps. pending beyond 45 work. days	358	91	688	4,516
No ACL assigned	0.13	0	0.33	4,516
Female ACL	0.34	0	0.47	3,947
Panel C: Applicant survey data				
Applicant age	47	47	13	2,760
Female	0.06	0	0.23	2,869
Applicant monthly income (BDT)	23,552	20,000	19,811	2,653
Applicant HH per capita expenditure (BDT)	4,396	3,400	3,579	2,869
Land value (BDT 100,000)	19	8	30	2,671
Land size (acre)	0.24	0.10	0.40	2,748
Typical payment amount (BDT)	6,718	5,000	8,416	1,802
Typical payment > 0	0.75	1	0.43	1,802
Reported payment amount (BDT)	1,477	0	3,480	2,869
Reported payment > 0	0.28	0	0.45	2,869

Notes: The table reports summary statistics for applications in the administrative data, offices, and applicants in the survey data. Observations in Panels A and C are inversely weighted by the number of observations in that 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 how much it is "normal for a person like yourself to pay." See discussion in Section 2.4.

Table 2: Scorecards' Effect on Application Processing Times

	(1)	(2)	(3)
	$\leq$45 working days	ln(working days)	Time index
Scorecard	0.060**	-0.126**	0.131**
	(0.027)	(0.059)	(0.059)
Start-month FE	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes
Observations	1,034,688	1,034,688	1,034,688
Clusters	311	311	311
Control mean	0.56	63.87	-0.00
Fraction imputed		0.02	

Notes: The table reports the effect of the scorecards on the speed of application processing. Column (1) shows the effect on the percentage of applications processed within the 45-working-day time limit. Column (2) shows the effect on the log of processing time. Applications that are not yet processed at the time of the final data transfer (December 2020) are given imputed processing times equal to the mean of processing times that are longer than the application has been pending. Column (3) shows the effect on an index combining the two outcome variables. The index is normalized to have a mean of zero and a standard deviation of one in the control group. The data contains all applications made between one month before the start of the experiment and 45 working days before the experiment ended (from August 13, 2018 to January 20, 2020). Standard errors are clustered at the land-office level. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.1.

Table 3: Scorecards' Effect on Bribe Payments

	Amount		Any Bribe		Amount If > 0	
	(1)	(2)	(3)	(4)	(5)	(6)
Scorecard	940	297	-0.018	-0.002	1,491*	1,170**
	(616)	(182)	(0.023)	(.022)	(763)	(453)
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	1,802	2,869	1,802	2,869	1,324	779
Clusters	112	112	112	112	112	111
Control mean	6,127	1,284	0.755	0.275	8,119	4,660
Bribe measure	Typical	Reported	Typical	Reported	Typical	Reported

Notes: The table reports the effect of the scorecards on bribe payments made for application processing. Columns (1), (3), and (5) show the effect of the scorecards on the response to the question about the value of a typical payment by "a person like yourself". Columns (2), (4), and (6) show the effect on the response to the question about actual payments to government officials or agents. Columns (3) and (4) show the effect on the percentage of nonzero responses to the questions (extensive margin). Columns (5) and (6) show the effect among applicants who reported a nonzero bribe (intensive margin). 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 are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.2.

Table 4: Scorecards' Effects by Office Baseline Performance

	Time index	Bribe payment	
	(1)	(2)	(3)
Scorecard × Overperform	0.034	2,069***	630***
	(0.080)	(765)	(233)
Scorecard × Underperform	0.238***	-100	42
	(0.088)	(958)	(259)
Overperform baseline	0.375***	-1833*	-815***
	(0.108)	(977)	(294)
P-value subgroup diff.	0.09	0.09	0.10
Start-month FE	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes
Observations	1,034,688	1,802	2,869
Clusters	311	112	112
Overperformers: q-value	0.51	0.02	0.02
Underperformers: q-value	0.01	0.85	0.77
Overperformers: control mean	0.27	5,455	944
Underperformers: control mean	-0.30	6,726	1,578
Bribe measure		Typical	Reported

Notes: The table reports the effect of the scorecards separately for offices with above- and below-median baseline performance. Column (1) is based on administrative data. Columns (2) and (3) are based on survey data. Column (1) shows the effects on the time index constructed from the two variables: whether the application was processed on time and the log of the overall processing time. Column (2) shows the effects on the response to the question of how much it is "normal for a person like yourself to pay" beyond the official fee. Column (3) shows the effects on reported payments to government officials or agents. The outcome variables in columns (2) and (3) 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 two hypotheses that the effect of the scorecards is zero for both overperformers and underperformers (Benjamini et al., 2006; Anderson, 2008). Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.3.

Appendix for Online Publication

A Theory

A.1 Model of Reputation and Bureaucrat Behavior

In this Appendix I present the formal model described in Section 6.

A.1.1 Model Setup

I assume that bureaucrats have the following utility function:

$$U(E_i, B_i) = D(E_i) + M(B_i) + t_i R(B_i, v_i P(E_i)) \tag{3}$$

where subscript i represents an individual bureaucrat. Disutility of effort $D(E)$ is a strictly concave function decreasing in effort E. The utility of bribe money $M(B)$ is a strictly concave function increasing in bribes B. $R(B, vP(E))$ is the utility the bureaucrat receives from reputation or pride in her work; it depends on B and visible performance $vP(E)$. The performance function $P(E)$ is concave and increasing in effort E. v is the visibility of performance; I assume v increases with the scorecards. $R(.)$ is a strictly concave function decreasing in bribes ($R_1(B, vP(E)) < 0$) and increasing in visible performance ($R_2(B, vP(E)) > 0$).

I assume that the only innate difference between bureaucrats is the weight t put on the reputation term in their utility function; this is what generates over- and underperformers in the model. It is not essential for the results of the model that the differences in performance are generated by differences in t; it could also have been generated by differences in ability or some other factor.

An important assumption is that bribes and visible performance are complements in generating reputation, i.e., $R_{12}(B, vP(E)) > 0$. In other words, honesty (the lack of bribes) and performance are substitutes. The logic for this assumption in terms of the bureaucrats' career concerns is that very honest bureaucrats can "get away" with being low-performing, and that high-performing bureaucrats can get away with being corrupt, but if a bureaucrat is both corrupt and a low-performer, this can have serious career consequences. Furthermore, if bureaucrats are both honest and high-performers, there is no way for them to accelerate their career path, because promotions are mostly determined by a bureaucrat's tenure in the bureaucracy. A more behavioral interpretation of this assumption is that it is a case of moral licensing, i.e., that bureaucrats feel a strong sense of shame if they are being both corrupt and low-performing, while if their visual performance improves they feel "licensed" to take more bribes without shame, and vice versa.

Another important assumption is that $P(E)$ can be both positive and negative. Specifically, I assume that $P(E)$ is positive when E is above median effort and negative when E is below median effort. This is realistic when performance is measured in a relative way, as it is in the scorecards' percentile rankings. This means that for bureaucrats performing above the median, increased visibility improves their visible performance, while for those performing below the median, it decreases visible performance.

As discussed in Section 6.1, I do not formally model applicants' behavior, but I assume that they have no choice but to accept the bureaucrats' bribe requests.

A.1.2 Solution to the Bureaucrats' Problem

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

$$D'(E_i^*) + t_i R_2(B_i^*, v_i P(E_i^*)) v_i P'(E_i^*) = 0 \tag{4}$$

where E_i^* and B_i^* represent the choices of E and B that maximize utility for bureaucrat i:

$$M'(B_i^*) + t_i R_1(B_i^*, v_i P(E_i^*)) = 0 \tag{5}$$

At the optimum, the marginal disutility of effort equals the marginal utility of effort's effect on reputational concerns (FOC 1), and the marginal utility of bribe money equals the marginal disutility from a decrease in reputation due to bribes (FOC 2).

A.1.3 Effect of Scorecards on Effort

Differentiating the first-order conditions with respect to v and solving for $\frac{\partial E_i^*}{\partial v}$ provides the expression in Equation 6. For ease of exposition, I henceforth drop the parentheses describing the variables that each function depends on, the subscript i, and the star superscript. All mentions of E and B refer to the values at the optimum.

$$\frac{\partial E}{\partial v} = P' \frac{-R_2 \left(M'' + tR_{11}\right) + vP\left(t\left(\left(R_{12}\right)^2 - R_{11}R_{22}\right) - M''R_{22}\right)}{\left(M'' + tR_{11}\right)\left(\frac{D''}{t} + R_{22}\left(vP'\right)^2 + R_2 vP''\right) - t\left(R_{12}vP'\right)^2} \tag{6}$$

Using the assumptions that $M(.)$,$D(.)$, and $R(.)$ are concave, the denominator is always positive. The first term in the numerator of Equation 6 reflects the incentive effect of the changed visibility of effort. I.e., it is the effect the scorecards have though changing the visibility of performance, holding the reputation level constant. This effect is always positive and is similar to the substitution effect during a change in wages in a labor-supply model. The second term in the numerator represents the reputation-level effect. I.e., the effect on effort stemming from the changed reputation level due to the change in visibility of the performance that was there prior to the scorecards. If $P < 0$, then this term is also positive, making the whole expression positive.[40]

Prediction 1: If $P(E) < 0$ then $\frac{dE}{dv} > 0$

[40] To define the incentive and reputation-level effects formally, I introduce two types of visibility, $\underline{v}$, which determines the level of reputation, and $\bar{v}$, which determines the incentive to improve performance. I then rewrite the FOCs as:

$$D'(E) + tR_2\left(B, \underline{v}P(E)\right)\bar{v}P'(E) = 0$$

$$M'(B) + tR_1\left(B, \underline{v}P(E)\right) = 0$$

The incentive effect is:

$$\frac{\partial E^*}{\partial \bar{v}} = P' \frac{-R_2\left(M'' + tR_{11}\right)}{\left(M'' + tR_{11}\right)\left(\frac{D''}{t} + R_{22}\left(vP'\right)^2 + R_2 vP''\right) - t\left(R_{12}vP'\right)^2} > 0$$

The reputation-level effect is:

The overall effect is ambiguous when $P(E) > 0$ since the reputational level effect is negative and may be larger or smaller than the incentive effect. The ambiguous effect is analogous to the effect of a wage increase on labor supply. The income effect and substitution effect go in different directions and depending on which dominates, the overall effect may be positive or negative.

A.1.4 Effect of the Scorecards on Bribes

Differentiating the first-order conditions with respect to v and solving for $\frac{\partial B_i^*}{\partial v}$ gives us the following expression:

$$\frac{\partial B}{\partial v} = R_{12} \frac{tvR_2 (P')^2 - P(D''(E) + tvR_2P'')}{(M'' + tR_{11}) \left(\frac{D''}{t} + R_{22} (vP')^2 + R_2vP'' \right) - t(R_{12}vP')^2} \tag{7}$$

The first term in the numerator is the impact from the incentive effect increasing effort. This leads to an improvement in visible performance, which in turn leads to an increase in bribes because of the complementary between visible performance and bribes. The second term is the impact of the reputation-level effect, which also affects bribes due to the complementarity between visible performance and bribes.

For bureaucrats with an effort above the median effort ($P(E) > 0$) the overall effect is positive. Their positive performance becomes more visible, so visible performance increases 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 lead 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 $\frac{dB}{dv} > 0$

$$\frac{\partial E^*}{\partial \underline{v}} = P' \frac{vP\left(t\left((R_{12})^2 - R_{11}R_{22}\right) - M''R_{22}\right)}{(M'' + tR_{11}) \left(\frac{D''}{t} + R_{22} (vP')^2 + R_2vP'' \right) - t(R_{12}vP')^2}$$

A.1.5 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. This does not qualitatively alter the directions of the predicted effects. 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 a reputation-level effect on all bureaucrats. Since the reputation-level effect is positive for effort and negative for bribes, in the model this would induce more effort and smaller bribe payments than the partial experimental rollout of the scorecards.

A.2 Monopolistic Price Discrimination Model

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 preanalysis 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 is the monopolist (Mussa and Rosen, 1978). Applicants get utility from having their application processed; the faster the application is processed, the more utility the processing generates. Applicants differ only in their willingness to pay for the speed of processing their applications. 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 are 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 applications faster, although this is not necessary for the main conclusions of the model.

Perfect information means that the bureaucrat can perfectly observe the applicants' willingness to pay for having the application processed within a certain time. 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. Under asymmetric information, only the applicants with the highest willingness to pay get their application processed at the Pareto optimal speed. All other applicants have their applications slowed down, as the bureaucrats trade-off providing faster processing for applicants with lower willingness to pay with how large a bribe they can charge from applicants with a higher willingness to pay.

A simple example is 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 for fast processing. 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 is therefore 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 Model's Predictions for the Experiment

The scorecards encourage bureaucrats to process applications within 45 working days. Section 4 shows that the scorecards led to an increase in the number of applications processed within 45 working days and that the effect was driven by offices that were underperforming 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 were processed faster. This is because for these applications, the value of the processing increased and they were 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 the option to pay less has increased.

A.2.2 Testing the Model's Predictions

The main testable prediction of this model under asymmetric information is that when processing times are improved, bribe payments should decrease among those who are getting their application processed quickly. Appendix Table A5 shows that the results from the experiment are inconsistent with this prediction. Column (1) shows the effect on bribes among those who had their applications processed quickly, using 25 working days as the cutoff. Among these applicants, the scorecards increased bribe payments by BDT 501. Column (2) shows that for applications processed outside of 25 working days, the estimated effect was a BDT 291 increase, but this effect is not statistically significant. Column (3) shows that even for offices that were underperforming at baseline, the effect of the scorecards on bribes for applications processed within 25 working days is estimated to be an increase of BDT 363. Although the effect is not statistically significantly different from zero, a large 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 applicants' willingness to pay for processing speed. However, even under full information, the bribe payments are not predicted to increase for those with the highest willingness to pay, and not among the overperforming offices that did not change processing times. The effects of the scorecards on bribes shown in Appendix Table A5 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. The information treatment is designed to alleviate this problem. However, column (4) of Appendix Table A5 shows that even for applicants who received the information in offices that were underperforming at baseline, no negative effect on bribes can be found for applications processed within 25 working days. Taken together, the results in Appendix Table A5 rejects the model's predictions.

B Additional Details on the Experiment and Data

B.1 Details on Randomization of Scorecards Treatment Assignment

The randomization in which land offices were assigned to receive the scorecards was done at the office level and implemented by the author. 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 randomization strata were created using the following variables:

- Number of applications processed within 45 working days in 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, strata were created based on offices being below the first tertile, between the first and second tertiles, or above the second tertile in the distribution of these variables. Since the group of offices with full e-governance implementation was smaller, strata were created based on offices being above or below the median in the distribution of these variables. The original randomization contained 114 land offices, two of which turned out not to be connected to the e-governance systems. The administrative data from these offices had been generated for training purposes and is excluded from the analyses.

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:

- Position relative to the first and second tertile in terms of number of applications processed within 45 working days in March 2019
- Position relative to the first and second tertile in terms of applications pending for more than 45 working days at the end of March 2019

Within each stratum, half of the land offices were randomly assigned to treatment.[41] If there was an odd number of offices in a stratum, I grouped the last office 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, I grouped these 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.

The peer performance list was randomized in the same way as the second wave of scorecard randomizations using data as of September 2019.The randomization only took place among the 155 land offices receiving the scorecards.

B.2 Details on Randomization of Information Intervention to Applicants

The information intervention was assigned to all applicants taking the in-person survey on randomly selected land-office survey days. The enumerators were informed about whether they should deliver the information treatment or not using printed schedules for the first two weeks of surveying for each land office. If the survey was conducted over

[41] The random assignment was implemented by the author using the Stata command runiform.

more than two weeks, the schedule should be followed from the start again. Weekdays were randomly selected to have the information delivered. The randomization was stratified by the weekday pairs: Sunday-Monday, separately for the first and second survey week, Tuesday-Wednesday, separately for the first and second survey week, Thursdays of the first and second survey week together. No surveying was done on the Friday and Saturday weekend. Noncompliance with the treatment assignment is discussed in Appendix C.4.1.

B.3 Data

B.3.1 Descriptive Statistics for Performance Indicators

The two performance indicators were chosen after discussions with both the Ministry of Land and the Access to Information (a2i) agency of the Government of Bangladesh. The first indicator is the number of applications processed within 45 working days in the past month, and the second indicator is the number of applications pending beyond 45 working days at the end of the month. The performance scorecards are based on all applications made in the land office, regardless of which ACL was assigned to the office when the application was made. Percentile rankings are calculated using the performance indicators for all the offices in the experiment, not just those in the treatment group. If several offices had the same value of the indicator, they both received the highest percentile in the percentile range covered by the offices. This caused offices having zero applications processed within 45 working days receiving a nonzero percentile ranking, and several offices that had zero applications pending received a 100th-percentile ranking, thus the average percentile ranking is above 50 and can vary over time. A thumbs-up symbol is shown next to the percentile ranking for all rankings above the 60th percentile while a thumbs-down symbol is shown next to all rankings below the 40th percentile.

Both indicators are absolute and not relative to the total number of applications processed or received. I chose to base the performance indicators on absolute numbers for two reasons. First, a relative number could have created incentives to not let applicants apply, and then focus on processing the applications of those allowed to apply. Second, a measure relative to the number of applications received was perceived by the government as being unfair toward larger land offices, where it may be harder to keep a high share of applications processed on time, due to a larger workload. The two absolute measures created a more-even playing field, as one measure (the number of applications processed on time) benefited larger land offices, while the other (the number of applications pending beyond the time limit) benefited smaller land offices.

Appendix Figure A9 shows that this worked in the intended way. The percentile ranking for the number of applications processed within the 45-working-day time limit and the number of applications pending beyond the 45-working-day time limit are correlated with the number of applications received in the land office in that month. However, because one correlation is positive and the other negative, the average percentile ranking does not have a strong correlation with the number of application received.

An important question regarding the performance scores is the extent to which they are controlled by the bureaucrats and the extent to which they are predetermined by the characteristics of a particular land office or other factors outside the control of the bureaucrat, such as the time of the year. I address this question using the adjusted R-squared from fixed-effects regressions where the monthly average performance ranking is the outcome variable, for land offices that had more than one bureaucrat during the span of my data. Including fixed effects for the land office and month yields an adjusted R-squared of 0.44 suggesting that less than half of the variation in the performance score is determined by land office characteristics and seasonality. Including bureaucrat fixed effects increases the adjusted R-squared to 0.63. This suggests that bureaucrats have a significant level of control over their performance score and that there is variation in bureaucrat performance that can explain a substantial part of the variation in performance scores.

B.3.2 Administrative Data From the E-governance System

The government partner transferred the updated administrative data at the beginning of each month from August 2018 until December 2020. The administrative data is at the application level and includes all applications made in the e-governance system since its inception in February 2017; due to privacy concerns only data without personal identifying information was shared.

To calculate the number of working days between the date the application started until the date it was processed, I use data on public, national, and general holidays in Bangladesh to construct a calendar consisting only of working days.[42] I then convert the application date and processing date to the working days calendar, assigning any date that is not a working day to the next working day. I then take the difference in working days between the application and processing date in the working day calendar. For the very small number of applications that were processed within the same day working day that the applications were made, I use the the exact time of the application and processing to create a measure of what fraction of a day it took to process the application.

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 the bureaucrats do not have access to. While it would be possible to create fake applications in the e-governance system, to process these applications with an acceptance, the processing fee would have to be paid. Creating fake unprocessed applications would decrease, not increase, a bureaucrat's performance ranking.

Some observations were deleted from the administrative data because they were not real applications, but rather applications that had been made for training purposes. During the training of bureaucrats in using the e-governance system, several example applications were made in two land offices that had not yet installed the e-governance system. 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. Thus, these offices were included in the

[42] The data was retrieved from https://www.timeanddate.com/holidays/bangladesh/.

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. I also found out that some other applications in the e-governance system are 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.

The administrative data contains the size of the land that the land-record change was made for. Some of the values in this variable are unrealistically high, most likely due to mistakes in what unit was used when inputting the land value. I clean this variable by setting any observation with a value of above 10 acres to missing as land transactions above this threshold are very rare and most likely the result of incorrect data entries. The largest land-record change in the survey data is 7.3 acres.

B.3.3 Administrative Data on ACL Assignments

For the analysis, 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 to land-record change applications that 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 an ACL was assigned to that office both prior to and after that month, in which case I conclude that the ACL was assigned to that office without making any updates in that month.

B.3.4 Survey Data

The survey was carried out in two stages: in the first stage, all individuals approaching the land office were asked if they were there for the purpose of a land-record change and if they consented to being part of the survey. A total of 3,213 applicants were interviewed in these in-person interviews. The follow-up interview was conducted by phone on the phone number provided by the applicant in the in-person interview. Phone numbers were contacted in the same order as the in-person interviews were conducted. Enumerators attempted to reach applicants three times in a day, and then made another three attempts the next day. If all attempts failed, the phone number was kept for another round of attempts one to two months later. Out of 3,213 in-person surveys 89% were successfully interviewed in the follow-up survey, yielding a final sample of 2,869 respondents. The average time between the two interviews was 3.3 months. 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.

Two types of duplicate observations are excluded from the survey data. One comes from the same person being interviewed more than once, probably as a result of them visiting the land office multiple times during the survey period, or multiple applicants providing the same phone number to be contacted on for the follow-up survey, probably because these individuals were in the same households. These observations may refer to different applications for land-record changes by the same individual or individuals in the same household, but since it was not possible to distinguish between the applications in the phone surveys, I keep only the first follow-up phone survey for each phone number. The main results are similar when including multiple follow-up surveys from the same phone number.

The other type of duplicate stems from multiple calls being made as follow-ups to the same in-person interview. This is natural, since many applicants were not reached on the first attempt. However, in some cases there were multiple interviews where the applicant started providing answers to interview questions but where the interview was redone, either because it was interrupted or for other reasons. I keep only the data from the final follow-up interview with a complete set of answers for each in-person interview.

Processing times in the survey data for applicants for whom the application was not processed at the time of the follow-up interview are imputed using the procedure described in Section 2.4.1. For applicants who did not answer the question about their monthly income, or provided an answer below BDT 1000, I impute their income using the income predicted from a regression of income on per capita household expenditure.

B.3.5 Measuring Bribes

Corruption is notoriously difficult to measure precisely (Olken and Pande, 2012). Due to the sensitive nature of the questions, I only asked applicants about bribes in the phone interview, not when interviewing them outside the land office. I also asked about bribe payments in two different ways, designed to elicit an answer even from those who were not comfortable talking about their own payments. The first question about bribe payments was phrased as "How much do you think it is normal for a person like yourself to pay in order to get the mutation processed and receive the khatian? Include all extra payments or gifts to agents, government officials, and other individuals, but do not include the 1150 taka official fee." This is the question I refer to as the "typical payment" and it is my preferred measure of bribe payments as it has a smaller percentage responses being zero suggesting that more respondents were comfortable answering the question. The next question was phrased as, "Did you pay any fees or give any gifts to anyone working for the Upazila Land Office or Union Parishad Land Office?" if the applicant responded yes, the enumerator went on to ask about to whom the payment was made to and "How much taka did you pay or what was the monetary value of the gift that you paid to [recipient]?" The "reported payment" outcome variable is the sum of all such reported payments.

To validate the magnitude of my bribe estimates in Appendix Figure A4, I compare them to an independent estimate created by Transparency International Bangladesh (TIB) as part of their nationally representative National Household Survey (Transparency International Bangladesh, 2016). There are two main reasons for why this measure may be different from my estimate of the average bribe payments, other than random differences between samples. First, the survey was done in person, potentially allowing enumerators to build more rapport with the respondents. Second, the survey was done in a nationally representative sample and for the period between November 2014 and October 2015. In the

TIB survey, 605 of the households had made applications for land-record changes, among these 57% reported having paid a bribe. The bribes reported in the TIB survey are on average higher than the bribes reported in the phone survey I conducted, but that difference shrinks substantially when excluding respondents reporting zero bribes. The average typical bribe payment reported in my survey lies between the average payment reported in the TIB survey and the average nonzero response. Overall, it is reassuring that the two different measures are of similar magnitude, despite using different methodologies, covering different areas, and being done for different periods.

B.3.6 Attrition in the Survey

The overall attrition rate from the in-person interview to completing the phone follow-up survey was 11%. Appendix Table A16 provide the estimates for 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%. Column (2) shows that the effect of the scorecards on attrition is somewhat higher in overperforming offices. Columns (3) and (4) show that the information intervention did not affect attrition.

A different definition of attrition is if a respondent did not answer a specific question in the survey. This definition also starts with the applicants in the in-person interview, but classify applicants who did not provide an answer to a specific question as having attrited, regardless of if these applicants were interviewed for the follow-up survey or not. In columns (5) and (6) of Table A16 I show that when defining attrition in this way for the question about the typical bribe payment, the scorecard treatment did not affect attrition, neither overall nor for over- or underperformers.

If the scorecards caused some applicants to not take the follow-up survey 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 test, since the effects on bribe payments are positive (overall and for overperforming offices) or non-negative (for underperforming offices). If the estimated effect of the scorecards on attrition for a particular subgroup is positive, I create lower Lee bounds by adding back a random selection of the applicants from treated offices for whom there is no follow-up survey data. The number of observations added back is determined so that there is no longer a difference in the attrition rate between the treatment and control group. I then set the bribe payments for this sub-sample to be zero, since that is the lowest possible bribe payment. If the effect of the scorecards on attrition is instead negative, I add back randomly selected respondents to the control group and assign these with the maximum bribe reported in the whole sample. I then conduct the main analysis from columns (1) and (2) of Table 3 and columns (2) and (3) of Table 4. The results are shown in Appendix Table A17. The lower Lee bounds are not qualitatively different from the main results. The estimated effect of the scorecards on bribe payments is still positive and statistically significant for offices overperforming at baseline.

B.3.7 Comparing the Effects in the Administrative and Survey Data

Throughout the paper, I use administrative data to measure the effect of the scorecards on processing times. There are several reasons for why the administrative data on processing times is superior to the survey data. First, the administrative data has a larger number of observations and clusters. Second, the administrative data does not suffer from recollection bias. Third, as the follow-up survey was conducted three months after the in-person survey, 10% of the applications had not yet been processed at the time of the follow-up survey and thus processing times had to be imputed for these applicants.

In Appendix Table A18, I directly compare the results using the administrative and survey data to rule out some potential, but unlikely, problems with the administrative 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 be accessed only by a few government contractors. I worked closely with this group—there are no suggestions that they were ever contacted by bureaucrats to alter the data on the server. A similar concern in the survey data 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.

To make the results in the administrative and survey data comparable, I restrict both samples to applications made in the land offices that were surveyed and applications made before May 26, 2019, the last day of the survey.[43] I exclude any applications made before February 1, 2019, to focus on applications made during the time period that the survey focused on.[44] Columns (1) and (3) of Appendix Table A18 show the estimated effects using administrative data, while columns (2) and (4) show the results using survey data. Comparing the main results from Table A6 with the results using administrative data during the survey period in the offices that were part of the survey shows that the estimated effects are similar but that when excluding the majority of the data and clusters, the estimates become less precise and not statistically significant. Comparing the results from the administrative data and survey data shows that the effects are qualitatively similar both for the overall and heterogeneous effects, but that the magnitudes are somewhat smaller in the survey data. Hence, if the results were driven by manipulations of the data, the bureaucrats would have had to manipulate both the administrative and the survey data, which is exceedingly unlikely.

[43]While there are application numbers making it theoretically possible to match applications from the survey with applications in the administrative data, this is very difficult in practice. The e-governance system has one global identification number. Unfortunately, only a small share of applicants had their digital application ID available to share with the enumerators. There also exist several other ID numbers in the same process, such as a local serial number for the application, for the record of rights, and for the plot of land. This confused some applicants, causing them to report the wrong application number to the enumerator.

[44]83% of applications in the survey were made after February 1, 2019.

C Additional Empirical Analyses

C.1 Potential Bias From Applicant Survey and Information Intervention

A 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 were carried out both in offices receiving scorecards and in control offices, the only situation in which such an effect would have biased the estimates for the effect of the scorecards is if it changed the effect of the scorecards.

To rule out such bias, Appendix Table A19 conducts 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 A19 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.2 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 discuss four potential unintended consequences the performance scorecards could have led to.

C.2.1 Gaming the Performance Indicators

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 implementing 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 the percentage of applications. However, the number of applications pending beyond the time limit would still be easier to keep down with fewer applications.

Column (1) in Panel A of Appendix Table A20 shows that the scorecards did not substantially affect the number of applications received in the e-governance system. Column (1) in Panel B shows that the scorecards did not decrease the number of applications more in the offices that were underperforming at baseline and where the scorecards had the largest effect on processing times.

Another potential problem could be that bureaucrats allow applications only from individuals for whom they know it is easier to process the application within the time limit. The size of the land for which the land-record change is being made is positively associated with the processing time.[45] Therefore, if bureaucrats intended to avoid accepting complex applications, we would also expect to see a decrease in applications average land size. Column (2) in Panel A of Appendix Table A20 shows that the scorecards did not substantially affect the average land size among applications received. Column (2) in Panel B shows a negative point estimate for the effect of the scorecards on application land size in underperforming offices, but that effect is imprecisely measured and does not provide conclusive evidence of the heterogeneous effects.

C.2.2 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 accept or reject 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) in Panel A of Ap-

[45] In a simple regression of log processing time on log land size, controlling for application month and land-office fixed effects, the coefficient on land size is 0.012, statistically significant at the 1% level.

pendix Table A21 shows that the scorecards did not substantially change the percentage of applications rejected overall. Column (1) in Panel B shows that for overperforming offices, the point estimate for the effect is positive, while for underperforming offices the point estimate is negative. However, both coefficients are small and not statistically significant.

Even if the rejection rate did not change, it is still possible that the quality of the decisions was worse. If this was the case, both more applications that should have been accepted, were rejected, and more applications that should have been rejected, were accepted. If an application is wrongfully rejected, applicants typically reapply in the same office. Therefore, the percentage of applicants reapplying after having been previously rejected can be used as an indicator for the percentage of incorrect rejections. Column (2) in Panel A of Appendix Table A21 shows that the percentage of applicants stating that they were reapplying, after previously having been rejected, increased with the scorecards, but the estimate is imprecise and not statistically significant. Column (2) in Panel B shows that the increase is driven by an increase in overperforming offices, while there was no increase in underperforming offices.

Together, these results suggest that the scorecards did not lead to a decrease in the quality of decision making because bureaucrats were pressured to make faster decisions, since both rejections and incorrect rejections were unaffected by the scorecards in underperforming offices, the offices that improved their processing speed. Instead the results are consistent with applicants in overperforming offices not satisfying the new higher bribe demands in these offices and are therefore rejected. However, these results should be interpreted with caution as the low-rate of reapplying applicants causes the estimated effects to be imprecise relative to the control group mean.

C.2.3 Spillover Effects on Applications Not Measured by the Scorecards

Even after the e-governance system had been installed, not all applications were made in the e-governance system, as described in Section 2.1.4. In the survey data, 24% of applications were not made using the e-governance system and did therefore not count toward the performance indicators in the scorecards. If the bureaucrats had reacted to the scorecards by diverting attention away from any task that was not measured by the scorecards, we would expect a negative spillover from the scorecards on the processing times for these applications. In Appendix Table A22, I estimate the effects of the scorecards on the number of visits needed and processing times for applications made outside of the e-governance system. Due to the decreased sample size, the estimates are imprecise, but overall there is no evidence for negative spillovers. Instead, the point estimate for the effect on the ICW index of the three outcome variables is an improvement of 0.05 standard deviations. Just as for the applications made in the e-governance system, the improvement is driven by offices underperforming at baseline, which is consistent with positive spillovers. While these results do not rule out all potential negative spillovers from the scorecards on other tasks performed by the bureaucrats, it suggests that the bureaucrats did not "game the system" by focusing only on those applications that would improve their scorecards.

C.3 Alternative Explanations for the Effects of Scorecards

C.3.1 Increased Marginal Costs of Bureaucrats' Time or Increased Willingness to Pay Among Applicants

One potential explanation for the scorecards causing higher bribe payments is that scorecards increase the bureaucrats' workload and therefore also the opportunity cost of their time. If bribe payments are made for bureaucrats to spend time on a particular application, this could increase bribe payments. Another alternative explanation is that faster processing times cause applicants to be willing to pay more to get their land-record change. However, both of these explanations are inconsistent with the results that in the underperforming offices where the changes in processing times are the largest, bribe payments do not change. Instead, bribe payments increase in the overperforming offices where changes in processing times are small. Therefore, it is unlikely that either of these mechanisms are substantial drivers of the effect on bribe payments.

C.3.2 Bureaucrat Transfers

An alternative explanation that is consistent with the heterogeneity in the effects on processing times and bribes is that overperforming bureaucrats get transferred due to receiving positive scorecards and that they are replaced by average-performing bureaucrats. If the average-performing bureaucrats have both slower processing times and demand higher bribes, we expect that bribe payments would increase in offices overperforming at baseline. Processing times may not change, because the incentive effect of the scorecards may cancel out the effect of the bureaucrat transfers.

This explanation is refuted by the data. Appendix Table A15 shows that the scorecards did not affect bureaucrats' transfers. Column (1) shows the effect on the monthly probability of being transferred, and column (2) shows the heterogeneity in the effect by offices overperforming and underperforming 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.

C.3.3 Supervisors Demanding Higher Bribes

I cannot observe sharing of bribe money within the bureaucracy, but it is plausible that this happens (Sanchez De la Sierra et al., 2020; Bandiera et al., 2021). If the scorecards led to the supervisors of overperforming bureaucrats demanding that larger amounts be shared with them, this could drive up bribe payments. The most plausible reason for such a demand would be a mechanism similar to the one in the model. In other words, the model would be correctly describing the mechanism, but with the wrong person described as the "bureaucrat." This seems unlikely, since for the supervisors, processing land-record applications is a very small share of the government services they are responsible for, while for the ACLs, it represents around a third to half of their work. It is therefore more likely that the scorecards cause a change in the ACLs reputation level and incentives than those of the supervisors.

C.3.4 Bureaucrats Using Scorecards in Negotiations With Applicants

Another alternative explanation is that positive scorecards help bureaucrats prove to applicants that they can process applications quickly. This could then allow the bureaucrats receiving positive scorecards to charge higher bribes, while it would not affect bribes in offices receiving negative scorecards. This explanation is implausible for four reasons. First, the coefficients on "Overperform baseline" in columns (4) and (5) of Appendix Table A14, show that the expected processing times are 13% lower in the overperforming offices not receiving the scorecards, suggesting that the applicants are already aware of the faster processing times in these offices. Second, column (4) shows that the point estimate for the effect of the scorecards on applicants' expectations of processing times in overperforming offices is a 5% increase. Although the positive effect is not statistically significant, if the scorecards helped bureaucrats improve applicants' expectations of processing times, the effect on the expectations should be negative. Third, 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 bureaucrats and applicants was it ever mentioned that the scorecards were shown to applicants. Fourth, the information intervention tried to accomplish the effect that a bureaucrat could have achieved by showing the scorecards to applicants, but Appendix Table A3 shows that the information intervention did not substantially increase bribes.

C.4 Effects of Information Intervention on Bribes

Appendix Table A3 shows that the information treatment did not affect bribes, neither by itself nor in combination with the scorecards. Columns (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 the results separately for overperforming and underperforming land offices. In particular the coefficients on $Info \times Scorecard \times Underperform$ show that even for applicants receiving the information intervention in offices that were underperforming at the start of the experiment, where both actual processing times and the expectations of these processing times declined the most, bribes did not decrease. Columns (3) and (6) also show that the positive effect of the scorecards on bribes, among offices overperforming at baseline, is similar for applicants receiving the information intervention and applicants not receiving the information intervention.

C.4.1 Noncompliance With the Information Intervention's Treatment Assignment

Enumerators implemented the information intervention during the in-person interviews. They were given a different schedule for each land office where each weekday was indicated as a day when the treatment would be delivered or not. When a day was indicated as a treatment day, the enumerators were asked to provide the information leaflet and explain its content in the middle of the interview. After having given the leaflet to the applicant, the enumerator would indicate in the survey that the information had been given and move on to the next question. According to the enumerators' indications of whether the information was given, in 17% of interviews incorrect information was given. This was due to enumerators' misunderstanding the schedule. Among the interviews where the treatment assignment was not followed, 91% were on days when the median interview was not given the information as per the treatment assignment.

The main result related to the information intervention is that it did not decrease bribe payments. In particular, that it did not decrease bribe payments in offices receiving the scorecards, as discussed in Section 5.3.1. Since this is a null result, I make conservative choices regarding the empirical strategy such that the information intervention has the highest ex-ante probability to cause a decrease in bribe payments. In the paper's main analysis, I therefore use the median treatment delivered in a particular land office and survey day as the treatment variable.[46] Alternatively, I could use the treatment indicated by the enumerator for each individual applicant, but due to likely information spillovers within an office survey day, I prefer the median treatment of the office-survey day.

Appendix Table A23 shows the robustness of the estimated effects of the information treatment to using alternative definitions of the treatment variable. Column (1) uses my preferred treatment variable based on the median treatment delivered in the land-office survey day. Column (2) uses the individual treatment delivered for the applicant. Column (3) uses the treatment assigned, and column (4) uses the treatment assigned but drops observations where the treatment was not correctly implemented. The null result of the effect on bribe payments is robust across all treatment variable definitions, both overall (Panel A) and in the offices that received the scorecards (Panel B). In Panel C, I show the robustness of the estimated effect of the information intervention on the expected processing time. The estimated effect is the largest when using the median delivered treatment or the actual delivered treatment, confirming that this is the most relevant variable for the analysis of the effect on bribe payments. When using the assigned treatment the effect is small, but when limiting the sample to the applicants for whom the correct treatment was delivered, the effect is similar to the effect estimated using the delivered treatment. This is what we would expect to see if the information intervention had a negative effect on expected processing time. When using only the treatment assignment, the estimated effect is small, since many applicants in the control group received the treatment and reduced the difference between the treatment and control groups.

C.5 Estimating the number of processing days saved per year

To estimate the number of processing days saved per year, I multiply the estimates for the total number of processing days per year, had there been no scorecards, and the improvement in the processing time for the average application. The reduction in the number

[46] When there is no median, due to an equal number of information treatments and controls delivered or due to no reported treatments or controls in a land-office survey day, I use the treatment assigned instead of the median.

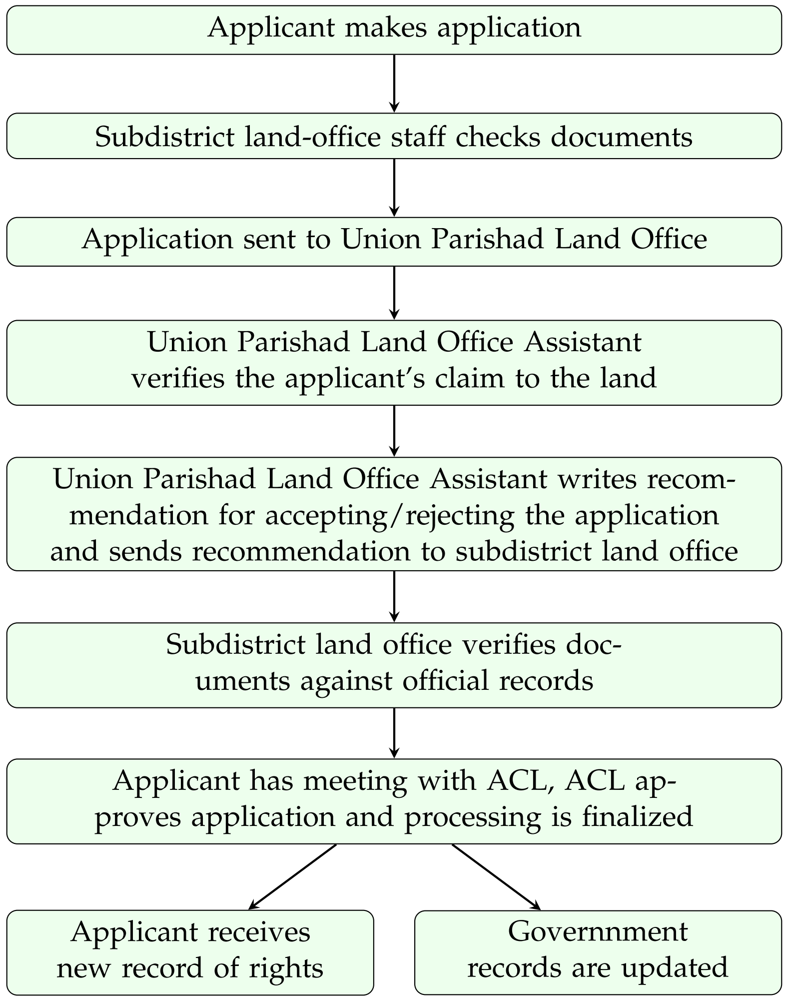
of processing days per year had there been no scorecards is estimated by taking the average processing time in the control offices and multiplying it by two times the number of applications in the treatment offices in the last six months of 2019, when the usage of the e-governance system had stabilized, but before the reduction in applications due to the COVID-19 pandemic. I then multiply this number with the percentage decrease for the average application, estimated by a uniformly weighted regression according to the specification in Equation 1.

C.6 Effects of the Scorecards on Applicant Satisfaction

Appendix Table A24 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 variable in regression Equation 1 and 2. Column (1) of Appendix Table A24 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 underperforming and offices overperforming at baseline. The negative effect is driven by offices that were overperforming at baseline, which is consistent with the observation that the scorecards increased bribe payments without improving processing times in these offices. Furthermore, despite that the scorecards were successful in reducing processing for offices underperforming 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 conclusions from the null result.

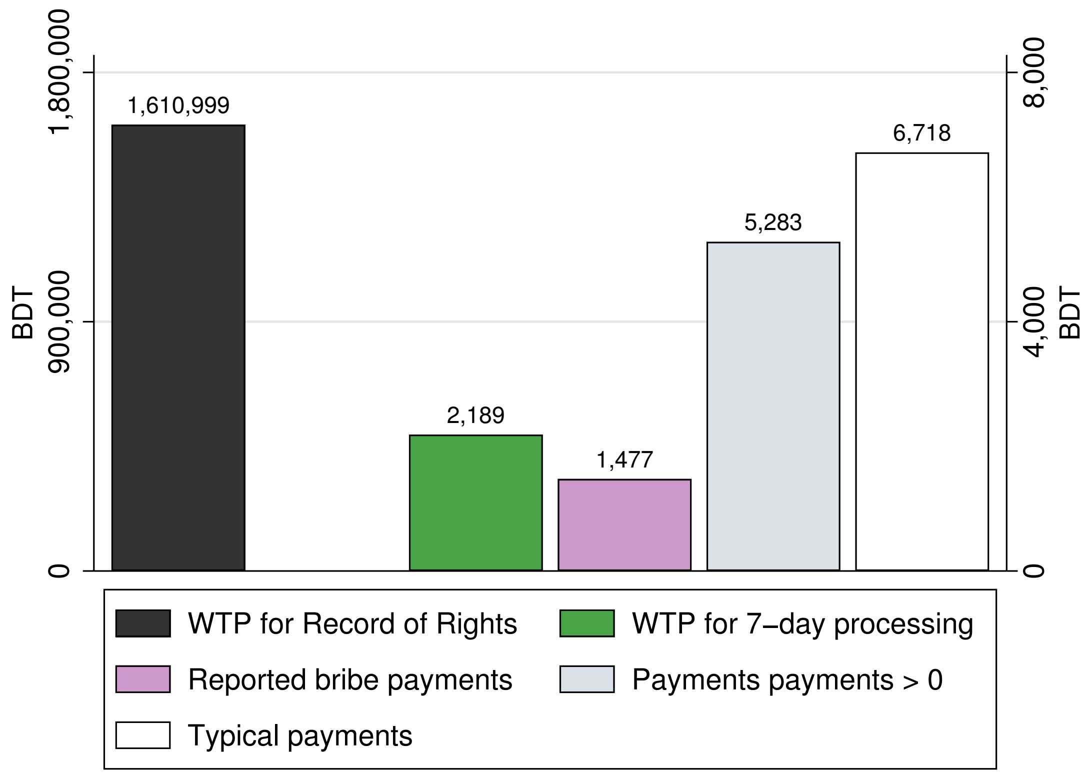
D Additional Tables and Figures

Figure A1: Application Process for Successful Application



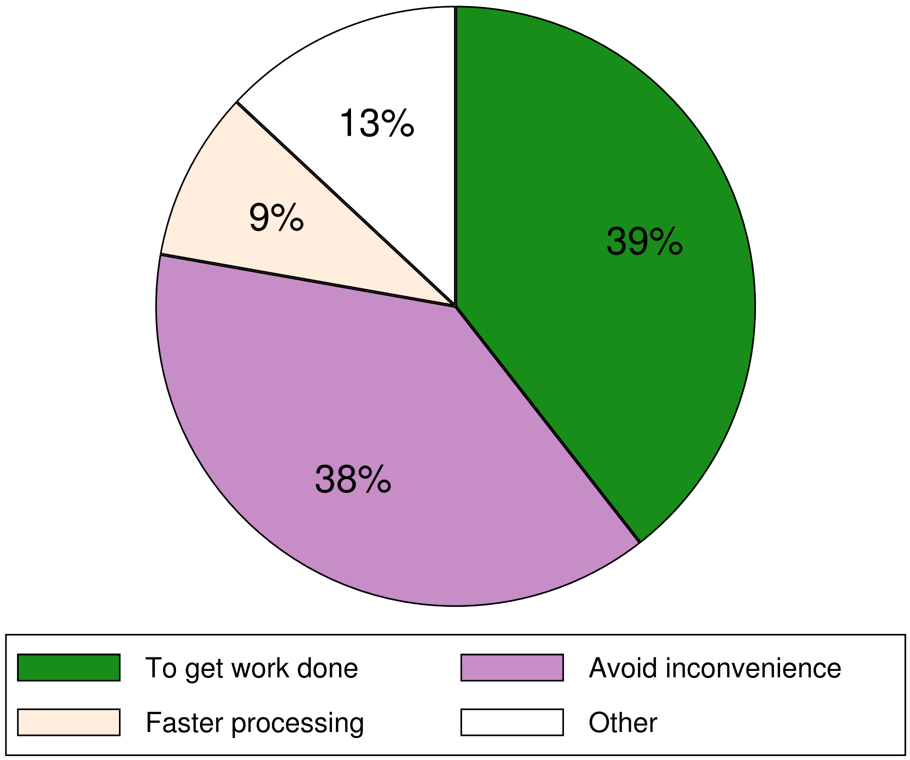
Notes: This figure depicts the *de jure* process for a successful land-record change application. See discussion in Section 2.1.

Figure A2: Value of Record of Rights, Faster Processing, and Bribes



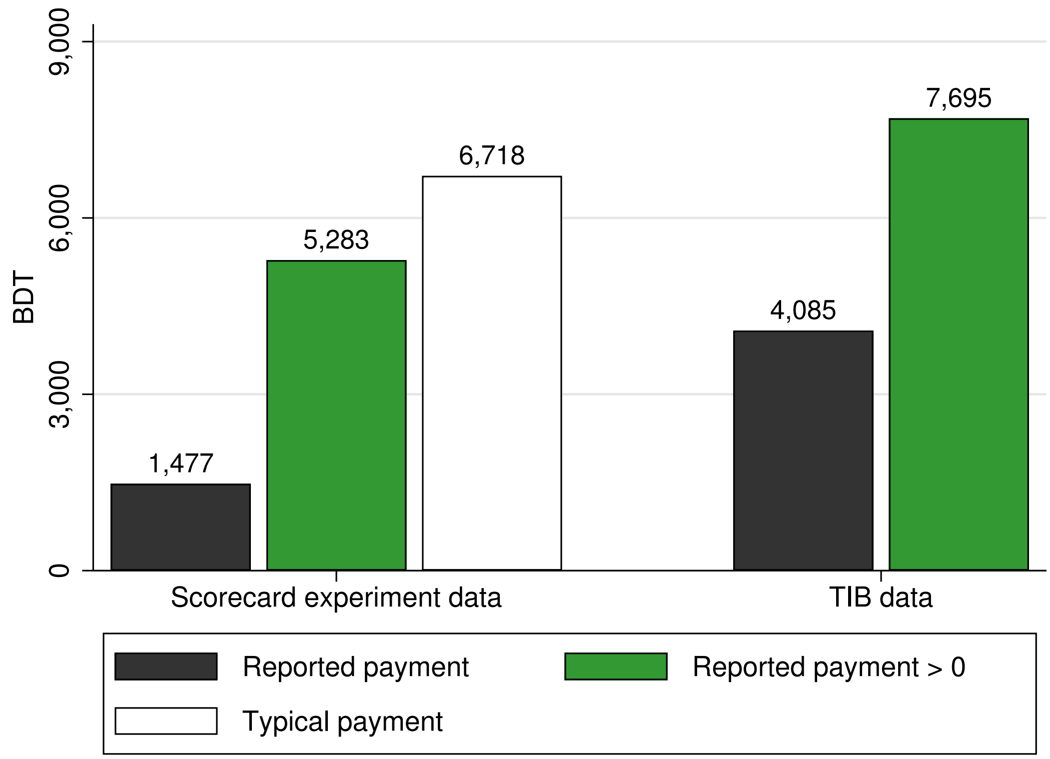
Notes: The figure displays the applicants' stated valuations of the land-record change and compare these to the bribe payments. The first bar shows the average of applicants' stated value of getting the land record change approved and receiving a record of rights. The second bar shows the average stated value of getting the application processed within seven days from the time of the first survey. The third bar shows the average value of bribe payments reported by the applicant; 73% of the applicants reported having paid no bribes. The fourth bar shows the average value of reported bribe payments among applicants reporting a nonzero bribe. The fifth bar shows the average response to the question about a typical bribe payment by "a person like yourself", 27% of the applicants responding to this question reported that a typical applicant paid no bribes. The first bar is measured on the left axis, the next four bars are measured on the right axis. All variables are winsorized at the 99th percentile. Observations are weighted by the inverse of the number of observations in that land office. USD/BDT≈84.3. See discussion in Sections 2.1 and 6.

Figure A3: Stated Reasons for Bribe Payments



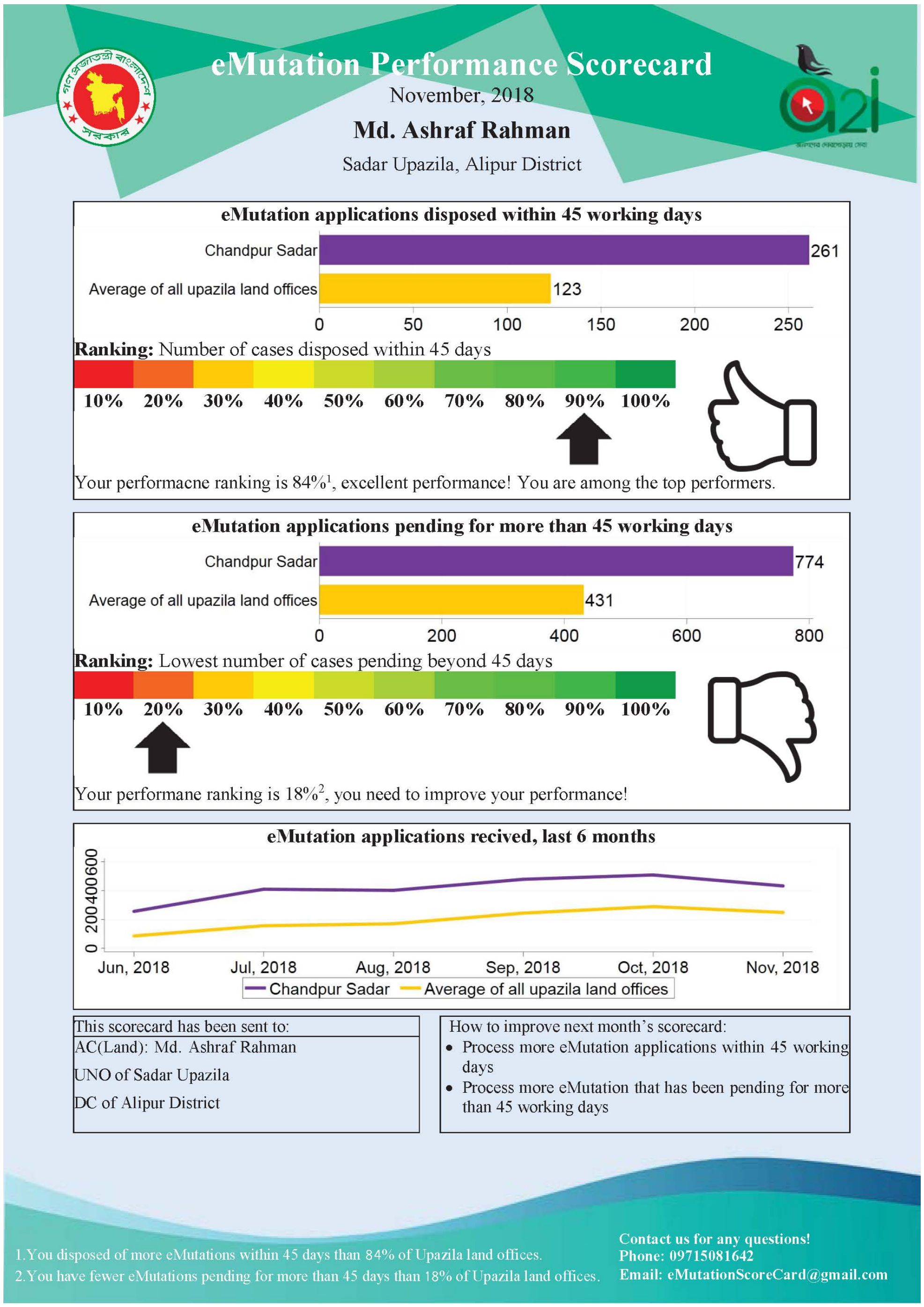
Notes: The figure displays the applicants' stated reasons for paying bribes. The responses are weighted by the amount of the bribe, so the percentages should be interpreted as the percentage of total bribes paid for that reason. The question was open-ended and was coded into response categories. See discussion in Sections 2.1 and 6.

Figure A4: Comparison of Estimated Bribes to the TIB Survey



Notes: The figure displays the average bribe payments reported in the phone survey conducted to evaluate the scorecard experiment and in an independent survey by Transparency International Bangladesh (TIB). The first bar shows the average value of bribe payments reported by applicants in the scorecard experiment phone survey; 73% of respondents reported having paid no bribes. The second bar shows the average value of bribe payments reported by applicants who reported having paid some bribe in the scorecard experiment phone survey. The third bar shows the average response to the question about the value of a typical bribe payment by "a person like yourself" in the scorecard experiment phone survey; 27% of respondents reported that a typical applicant paid no bribes. The fourth bar shows the average value of bribe payments reported by applicants in the TIB 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 who reported having paid some bribe in the TIB 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. USD/BDT≈84.3. See discussion in Appendix B.3.5.

Figure A5: Example of a Performance Scorecard



Notes: Example of a performance scorecard in English. The ACL name and land office name are changed to preserve anonymity. The English scorecard was accompanied by a scorecard in Bengali as well as an explanatory note showing how the numbers are calculated. See discussion in Section 2.2 and Appendix B.3.1.

Figure A6: Example of a Peer Performance List

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

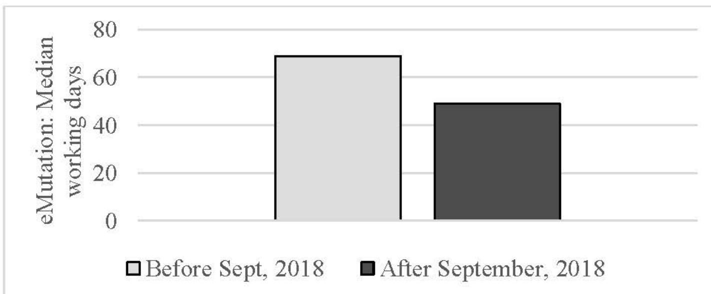
Upazila Land Office	District	Disposed ranking	Pending Ranking
[Name of upazila]	[Name of district]	43%	20%
[Name of upazila]	[Name of district]	48%	39%
[Name of upazila]	[Name of district]	51%	57%
[Name of upazila]	[Name of district]	17%	16%
[Name of upazila]	[Name of district]	18%	66%
[Name of upazila]	[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%

Notes: Example of the first page of a peer performance list; the full list contains two pages. The office and district names have been removed to preserve anonymity. See discussion in Section 2.3.1.

Figure A7: Leaflet Given to Applicants in the 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).

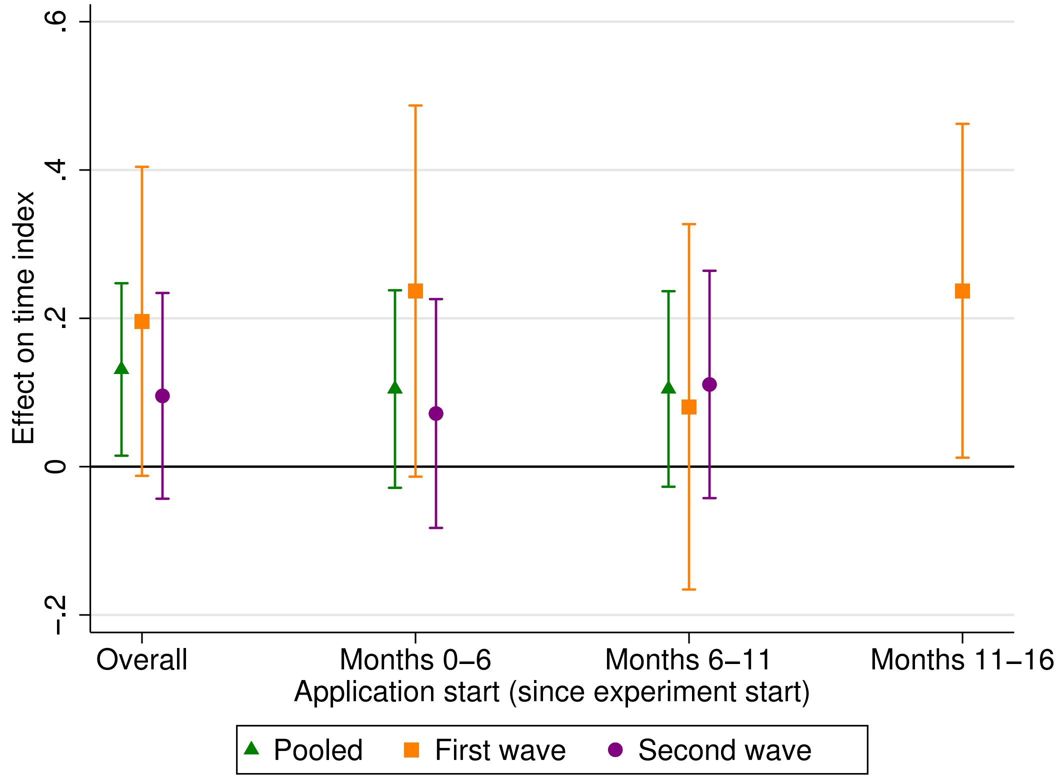
Steps of Land Record eMutation and timeline:

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]

Notes: English translation of the information leaflet given to applicants during the information intervention. The leaflet 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 leaflet was given to applicants in both treatment and control offices. See discussion in Section 2.5.

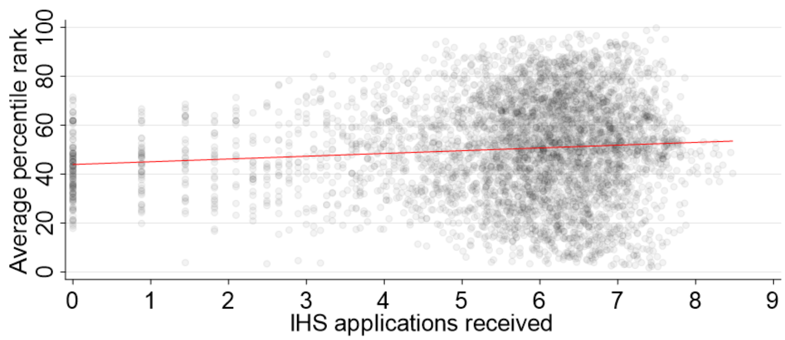
Figure A8: Scorecards' Effect by Time Since Start of the Experiment



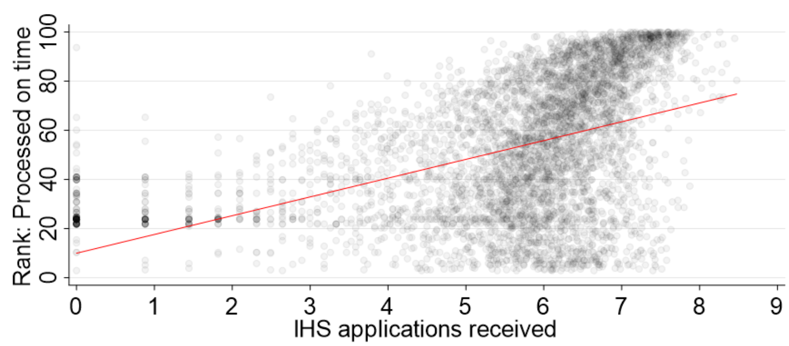
Notes: The figure reports the regression coefficients and confidence intervals for regressions using applications made during different periods relative to the start of the experiment. The effects are measured in standard deviations of the time index from column (3) of Table 2. The index is constructed using the variables for whether the application was processed on time and the log of the overall processing time. The estimate for each period comes from a separate regression restricting the sample to applications made in that period. The results are from regressions using data from all offices (triangles), offices in the first randomization wave (squares), and offices in the second randomization wave (circles). The months are numbered relative to when the first scorecards were sent out for that office's randomization wave. Month 0 is the month before the first scorecard. Confidence intervals are constructed using standard errors clustered at the land-office level. See discussion in Section 4.1.2.

Figure A9: Office by Month Performance and Land Office Size

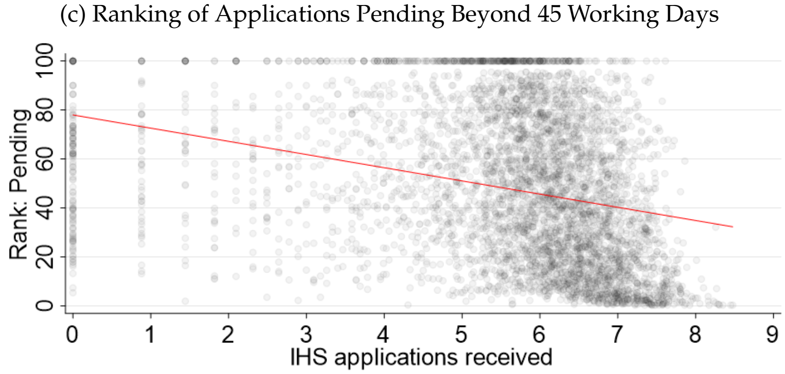
(a) Average Performance Ranking



(b) Ranking of Applications Processed Within 45 Working Days



(c) Ranking of Applications Pending Beyond 45 Working Days



Notes: The figure displays the relationship between the number of applications received by a land office in a month and the office's performance ranking in that month. The data comes from both treatment and control offices. Data is from when the office first started using the e-governance system until March 2020 (the end of the experiment). Sub-Figure a) shows the relationship between the average performance ranking in a month and the inverse hyperbolic sine (IHS) transformation of the number of applications received in that month. Sub-Figure b) shows the percentile ranking of the number of application processed on time and the IHS of the number of applications received. Sub-Figure c) shows the percentile ranking of the number of applications pending beyond the 45-working-day time limit and the number of applications received. See discussion in Appendix B.3.1.

Table A1: Balance of Randomization: Administrative Data

	Scorecard		Control		Difference
	Obs. (Cluster)	Mean (SD)	Obs. (Cluster)	Mean (SD)	Diff. (SE)
≤45 working days	56,703	0.50	56,564	0.53	-0.028
	(146)	(0.50)	(146)	(0.50)	(0.04)
ln(Process time)	56,703	3.77	56,564	3.69	0.083
	(146)	(1.40)	(146)	(1.44)	(0.13)
Time index	56,703	-0.06	56,564	0.00	-0.061
	(146)	(0.98)	(146)	(1.00)	(0.09)
Process time (w. days)	56,703	77.60	56,564	76.90	0.700
	(146)	(73.53)	(146)	(79.15)	(6.98)
Approved	36,409	0.70	38,776	0.73	-0.030
	(136)	(0.46)	(141)	(0.44)	(0.03)

Notes: The table reports the balance of randomization for treatment and control offices using administrative data from 45 working days before the first scorecard in the randomization wave was sent. Due to this restriction, only 292 of the 311 offices are part of the balance of randomization data. For applications not processed by the date the first scorecard was sent, the processing time is imputed using the procedure described in Section 2.4.1. Data on approvals are as per the date the first scorecard was sent. P-value for the F-test of joint orthogonality: 0.87. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 2.6.

Table A2: Balance of Randomization: Survey Data

	Scorecard		Control		Difference
	Obs. (Cluster)	Mean (SD)	Obs. (Cluster)	Mean (SD)	Diff. (SE)
Applicant age	1,383	47.34	1,377	47.39	-0.05
	(56)	(13.35)	(56)	(13.82)	(0.64)
Female	1,460	0.06	1,409	0.06	-0.00
	(56)	(0.23)	(56)	(0.24)	(0.01)
Monthly income (BDT 1K)	1,460	24.49	1,409	22.85	1.64
	(56)	(17.62)	(56)	(17.88)	(1.22)
App. status: Applying	1,382	0.23	1,377	0.22	0.01
	(56)	(0.42)	(56)	(0.42)	(0.03)
Ongoing	1,382	0.64	1,377	0.63	0.01
	(56)	(0.49)	(56)	(0.49)	(0.04)
Rejected	1,382	0.00	1,377	0.00	0.00
	(56)	(0.07)	(56)	(0.06)	(0.00)
Approved	1,382	0.13	1,377	0.15	-0.02
	(56)	(0.34)	(56)	(0.36)	(0.03)
Land value (BDT 100K)	1,335	20.70	1,336	17.32	3.38
	(56)	(31.87)	(56)	(29.82)	(2.17)
Land size (acre)	1,379	0.24	1,369	0.24	-0.00
	(56)	(0.41)	(56)	(0.39)	(0.02)

Notes: This table reports the balance of randomization for applicants in treatment and control offices using survey data. All data comes from the in-person survey of applicants, which was conducted before the conclusion of the processing of the applicants' applications but after the start of the scorecards. P-value for F-test of joint orthogonality: 0.90. Continuous variables are winsorized at the 99th percentile. Observations are inversely weighted by the number of observations in that land office. USD/BDT≈84.3. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 2.6.

Table A3: Effects of Information Treatment and Scorecards on Bribes

	Typical payment			Reported payment		
	(1)	(2)	(3)	(4)	(5)	(6)
Information treatment	212	-326	-311	-31	-56	-63
	(490)	(767)	(760)	(150)	(184)	(188)
Scorecard		462			285	
		(719)			(218)	
Scorecard x Information		987			29	
		(1,105)			(278)	
Info. x Scorecard x Overperform			1,644*			740***
			(926)			(280)
No info. x Scorecard x Overperform			2,427**			533*
			(956)			(280)
Info. x Scorecard x Underperform			1,343			-26
			(1,411)			(334)
No info. x Scorecard x Underperform			-1,519*			122
			(896)			(310)
Overperform baseline			-1,712*			-824***
			(983)			(294)
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	1,802	1,802	1,802	2,869	2,869	2,869
Clusters	539	112	112	570	112	112
Control mean	6,553	6,258		1,508	1,328	

Notes: This table reports the effect of the scorecard and information treatments on bribe payments. Columns (1)-(3) show the effect on the response to the question of how much it is "normal for a person like yourself to pay" beyond the official fee. Columns (4)-(6) show the effect on reported payments to government officials or agents. All outcome variables are winsorized at the 99th percentile. The outcome variables are in BDT. USD/BDT≈84.3. 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 observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Appendix 5.3.1.

Table A4: Effect of Peer Performance List

	≤45 working days		ln(working days)	
	(1)	(2)	(3)	(4)
Peer Performance List	0.002	-0.007	-0.030	0.037
	(0.045)	(0.048)	(0.094)	(0.105)
Post × Peer Performance List		0.008		-0.060
		(0.045)		(0.095)
Scorecard		0.048		-0.124
		(0.038)		(0.090)
Post × Scorecard		0.020		-0.002
		(0.038)		(0.083)
Start-month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	280,002	1,034,688	280,002	1,034,688
Clusters	155	311	155	311
Control mean	0.68	0.59	49.47	63.87

Notes: This table reports the effect of adding the peer performance list to the scorecard and sharing the performance information about one office with the ACLs, UNOs, and DCs of 76 other offices. Columns (1)-(2) show the effect on the percentage of applications processed within the 45-working-day time limit. Columns (3)-(4) show the effect on the log of processing time. Columns (1) and (3) only use data from offices receiving the scorecards and applications made one month before the first performance list was sent out until the end of the experiment (August 15, 2019 to January 20, 2020). Columns (2) and (4) use data on all applications made between one month before the start of the Scorecard experiment started and 45 working days before the experiment ended (August 13, 2018 to January 20, 2020). The indicator variable *Peer Performance List* takes the value of one for applications made in offices receiving the peer performance lists. The indicator variable *Post* takes the value of one for applications made later than one calendar month before the first performance lists were sent out. Standard errors are clustered at the land-office level. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 5.1.

Table A5: Testing Predictions From Monopolistic Price-Discrimination Model

	Reported payment			
	(1)	(2)	(3)	(4)
Scorecard	501**	291		
	(208)	(247)		
Scorecard x Overperform			633***	
			(234)	
Scorecard x Underperform			363	
			(363)	
Overperform baseline			-356	-304
			(337)	(336)
Info. x Scorecard x Underperform				19
				(396)
No info. x Scorecard x Underperform				814*
				(482)
Info. x Scorecard x Overperform				288
				(301)
No info. x Scorecard x Overperform				1,000**
				(392)
Information treatment				37
				(224)
Start-month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Sample	Appr. $\leq$ 25	Appr. $>$ 25	Appr. $\leq$ 25	Appr. $\leq$ 25
Observations	677	1,332	677	677
Clusters	111	111	111	111

Notes: This table reports the effects of the scorecard and information treatments on bribes for applicants with different processing speed. Column (1), (3), and (4) use data only from applications approved within 25 working days while column (2) uses data from applications that have an approval time of more than 25 working days. No rejected or still ongoing applications are included. USD/BDT≈84.3. Bribe amounts are winsorized at the 99th percentile. Standard errors are clustered at the land-office level. ***p<0.01; **p<0.05; *p<0.1. See discussion in Appendix A.2.

Table A6: Scorecards' Effects on Visits and Processing Times in Survey

Panel A: Overall effect	(1) Visits	(2) $\leq$45 w. days	(3) ln(w. days)	(4) ICW Index
Scorecard	-1.266***	0.040	-0.059	0.134**
	(0.422)	(0.030)	(0.049)	(0.058)
Panel B: Heterogeneous effects				
Scorecard $\times$ Overperform baseline	-1.054*	0.012	-0.005	0.083
	(0.606)	(0.049)	(0.078)	(0.088)
Scorecard $\times$ Underperform baseline	-1.272**	0.055	-0.088	0.155*
	(0.636)	(0.041)	(0.066)	(0.082)
Overperform baseline	-1.676**	0.113**	-0.216**	0.262**
	(0.784)	(0.051)	(0.084)	(0.105)
Start-month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	2,869	2,869	2,869	2,869
Clusters	112	112	112	112
Control mean	10.73	0.56	59.81	

Notes: This table reports the effect of the scorecards on visits to land offices and processing times measured in the survey data. Standard errors are clustered at the land-office level. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.4.

Table A7: Heterogeneous Effects on Processing Times, Controlling for Baseline Values Interacted With Treatment

	Time index						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Scorecard × Underperform baseline	0.204*	0.200*	0.180	0.197	0.219*	0.205*	0.184
	(0.121)	(0.120)	(0.121)	(0.121)	(0.119)	(0.121)	(0.119)
Scorecard × Rank: Applications within 45 working days before baseline		0.001					-0.004
		(0.002)					(0.003)
Scorecard × Rank: Applications prior to 45 working days before baselin			-0.002				-0.003
			(0.002)				(0.003)
Scorecard × Female ACL baseline				-0.039			-0.125
				(0.141)			(0.140)
Scorecard × No ACL baseline					-0.297*		-0.388**
					(0.154)		(0.158)
Scorecard × Land size (st. dev.)						-0.008	0.000
						(0.013)	(0.013)
Baseline control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline control		Yes	Yes	Yes	Yes	Yes	Yes
Start-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,034,688	1,034,688	1,034,688	1,034,688	1,034,688	1,007,014	1,007,014
Clusters	311	311	311	311	311	311	311

Notes: This table reports the heterogeneous effects on processing time by land office baseline performance, controlling for other variables interacted with the treatment. The first row of estimates shows the difference in the effect between under- and over-performing offices. Column (1) uses the main specification from Table 4. Columns (2) and (3) controls for the treatment interacted with two measures of office size: the offices' percentile ranking in terms of the number of applications received in the 45-working-day time period before the treatment and offices' percentile ranking in terms of the total number of applications received before that 45-working-day period. Columns (4)-(6) controls for the treatment interacted with the ACL being a female at baseline, no ACL being appointed at baseline, and the size of the land the application was for. Column (7) includes all the controls from columns (2)-(6). Standard errors clustered at the land-office level. Observations are inversely weighted by the number of observations in that land office. ***$p<0.01$; **$p<0.05$; *$p<0.1$. See discussion in Section 4.3.

Table A8: Heterogeneous Effects on Bribe Payments, Controlling for Baseline Values Interacted With Treatment

	Typical bribe payment						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Scorecard × Overperform baseline	2,169*	2,180*	2,594**	2,044	2,088	2,181*	2,691*
	(1,252)	(1,251)	(1,306)	(1,289)	(1,280)	(1,214)	(1,390)
Scorecard × Rank: Applications within 45 working days before baseline		-2					-35
		(22)					(29)
Scorecard × Rank: Applications prior to 45 working days before baselin			-20				-48
			(23)				(31)
Scorecard × Female baseline ACL				775			-38
				(1,508)			(1,589)
Scorecard × No baseline ACL					442		-125
					(1,708)		(1,758)
Scorecard × Land size (st. dev.)						113	107
						(379)	(382)
Treatment control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline control		Yes	Yes	Yes	Yes	Yes	Yes
Start-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,802	1,802	1,802	1,802	1,802	1,726	1,726
Clusters	112	112	112	112	112	112	112

Notes: This table reports the heterogeneous effects on bribe payments by land-office baseline performance, controlling for other variables interacted with the treatment. The first row of estimates shows the difference in the effect between under- and over-performing offices. Column (1) uses the main specification from Table 4. Columns (2) and (3) control for the treatment interacted with two measures of office size: the offices' percentile ranking in terms of the number of applications received in the 45-working-day time period before the treatment and offices' percentile ranking in terms of the total number of applications received before that 45-working-day period. Columns (4)-(6) control for the treatment interacted with the ACL being a female at baseline, no ACL being appointed at baseline, and the size of the land the application was for. Column (7) includes all the controls from columns (2)-(6). Standard errors are clustered at the land-office level. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.3.

Table A9: Scorecards' Effect on Processing Times: Alternative Specifications

	Time index					
Panel A: Overall effect	(1)	(2)	(3)	(4)	(5)	(6)
Scorecard	0.089	0.125*	0.140**	0.097	0.137**	
	(0.089)	(0.070)	(0.061)	(0.076)	(0.057)	
Scorecard × Post						0.156*
						(0.080)
Panel B: Heterogeneous effects						
Scorecard × Overperform	-0.009	0.001	0.005	0.031	0.032	
	(0.116)	(0.089)	(0.081)	(0.101)	(0.077)	
Scorecard × Underperform	0.197*	0.264***	0.280***	0.169	0.250***	
	(0.119)	(0.094)	(0.089)	(0.110)	(0.084)	
Overperform baseline	0.537***	0.572***	0.341***	0.487***	0.261**	
	(0.110)	(0.088)	(0.103)	(0.130)	(0.116)	
Scorecard × Post × Overperform						0.040
						(0.074)
Scorecard × Post × Underperform						0.283**
						(0.114)
P-value subgroup diff.	0.21	0.04	0.02	0.37	0.06	0.13
Start-month FE	No	No	No	Yes	Yes	Yes
Stratum FE	No	No	No	Yes	Yes	No
Weighted by office	No	Yes	Yes	No	Yes	Yes
Baseline controls	No	No	Yes	No	Yes	No
Office FE	No	No	No	No	No	Yes
Observations	1,034,688	1,034,688	1,034,688	1,034,688	1,034,688	1,188,351
Clusters	311	311	311	311	311	306

Notes: This table shows the robustness of the estimated effect of the scorecards on the time index to different regression specifications. The time index is constructed from the two variables: whether the application was processed on time and the log of the overall processing time. Panel A shows the estimates of the overall effect and Panel B shows the estimates of the heterogeneous effects, as in Tables 2 and 4, respectively. The specifications differ from the ones used in those tables in the following ways: column (1) shows the estimate from uniformly weighted regressions with no controls. Column (2) uses no controls. Column (3) uses no strata or month fixed effects, but it controls for baseline month values for the number of applications processed within 45 working days, the number of applications pending beyond 45 working days, the number of applications received, and the percentage of applications received in the month two months before the baseline that were processed within the time limit. Column (4) shows the estimate from a uniformly weighted regression. Column (5) controls for the baseline month controls. Column (6) shows the estimate from a regression, including applications made 45 working days before the start of the experiment (for the 306 offices that have such data) and using land-office fixed effects. Standard errors are clustered at the land-office level. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.5.

Table A10: Scorecards' Effect on Bribes: Alternative Specifications

	Typical payments						Reported payments						To gov. off.
Panel A: Overall effect	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Scorecard	1,194*	1,164*	875	688	1,890**	482	388*	386**	307*	238	610**	194	245*
	(717)	(698)	(623)	(522)	(853)	(449)	(196)	(195)	(172)	(179)	(264)	(127)	(141)
Panel B: Heterogeneous effects													
Scorecard x Overperform	1,808*	1,602*	2,267***	1,325**	2,811***	1,576***	647***	648***	659***	519**	1,011**	468***	462***
	(920)	(864)	(827)	(612)	(1,067)	(585)	(234)	(219)	(233)	(207)	(418)	(168)	(173)
Scorecard x Underperform	666	811	-393	-215	1,085	-534	175	175	28	75	255	-21	82
	(1,068)	(1,069)	(923)	(813)	(1,320)	(646)	(303)	(300)	(241)	(235)	(348)	(172)	(208)
Overperform baseline	-1,219	-1,272	-2,138**	-1,916**	-2,055*	-1,667**	-455*	-634**	-759***	-1,005***	-634	-636***	-565**
	(837)	(919)	(982)	(843)	(1,149)	(754)	(244)	(255)	(280)	(250)	(404)	(202)	(237)
P-value subgroup diff.	0.42	0.57	0.04	0.06	0.31	0.02	0.22	0.21	0.06	0.10	0.19	0.05	0.16
Start-month FE	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Stratum FE	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Weighted by office	No	Yes	No	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
PDS controls	No	No	No	Yes	No	No	No	No	No	Yes	No	No	No
Winsorized	99 pctl.	99 pctl.	99 pctl.	99 pctl.	No	95 pctl.	99 pctl.	99 pctl.	99 pctl.	99 pctl.	No	95 pctl.	99 pctl.
Observations	1,802	1,802	1,802	1,802	1,802	1,802	2,869	2,869	2,869	2,869	2,869	2,869	2,869
Clusters	112	112	112	112	112	112	112	112	112	112	112	112	112

Notes: This table shows the robustness of the estimated effect of the scorecards on bribes. Panel A shows the estimates of the overall effect and Panel B of the heterogeneous effects, similar to the estimates in columns (1) and (2) of Table 3 and columns (2) and (3) of Table 4. Columns (1)-(6) show the effect on the response to the question of how much it is "normal for a person like yourself to pay" in bribes. Columns (7)-(12) show the effect on reported bribe payments to government officials or agents. The specifications differ from the specifications in Tables 3 and 4 in the following ways: columns (1) and (7) use uniformly weighted regressions with no fixed effects. Columns (2) and (8) use regressions with no fixed effects. Columns (3) and (9) use uniformly weighted regressions. Columns (4) and (10) use post-double-selection (PDS) to select controls among 154 variables generated from the baseline administrative data and survey variables unlikely to have been affected by the experiment (e.g. applicant characteristics; land value, size, and usage; whether e-governance system was used; enumerator conducting the survey; and indicators for each quartile of continuous control variables). Columns (5) and (11) use unwinsorized outcome variables. Columns (6) and (12) use outcome variables winsorized at the 95th percentile. Column (13) uses payments made directly to government officials as the outcome variable, excluding payments made to agents. Standard errors are clustered at the land-office level. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.5.

Table A11: Scorecards' Effects 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	0.211*	-0.104	2.096	2.020	0.090
	(0.121)	(0.143)	(1.684)	(1.771)	(0.061)
Panel B: Heterogeneous effects					
Scorecard $\times$ Overperform	-0.022	0.171	-0.706	-0.845	-0.065
	(0.147)	(0.219)	(2.187)	(2.619)	(0.082)
Scorecard $\times$ Underperform	0.466**	-0.412**	5.171**	5.226**	0.263***
	(0.195)	(0.186)	(2.614)	(2.414)	(0.091)
Overperform baseline	0.393*	-0.422	6.368*	3.853	0.210*
	(0.226)	(0.280)	(3.321)	(3.539)	(0.125)
P-value subgroup diff.	0.05	0.05	0.09	0.10	0.01
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

Notes: This table reports the effect of the scorecards on office-by-month level outcomes. Panel A shows the estimates of the overall effect. Panel B shows the estimates of the heterogeneous effects. Column (1) shows the effect on the inverse hyperbolic sine (IHS) transformation 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 an ICW index created with the outcome variables of columns (1)-(4); a higher index value indicates a better performance. For the control group, the index has a mean of zero and a standard deviation of one. Standard errors are clustered at the land-office level. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.5.

Table A12: Scorecards' Effect on Processing Times: Alternative Functional Forms and Imputation Techniques

	$\leq$ 45 w. days	Working days		ln(w. days)	
	(1)	(2)	(3)	(4)	(5)
Scorecard	0.240*	-6.846*	-0.101*	-0.124**	-0.123**
	(0.134)	(4.120)	(0.0589)	(0.0581)	(0.0542)
Start month FE		Yes	Yes	Yes	Yes
Stratum FE		Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes
Observations	1,034,688	1,034,688	1,034,688	1,034,688	1,006,272
Clusters	311	311	311	311	311
Specification	Logit	OLS	Neg. Binomial		
Imputation				Office mean	Drop obs.

Notes: This table shows the robustness of the effect of the scorecards on processing times from Table 2 to different assumptions regarding the functional form of the relationship between the treatment and outcome variable and to the imputation procedure used to assign a processing time to the 2% of applications that are not yet processed. Column (1) uses a logit model to estimate the effect of the scorecards on the probability of a application being processed on time. Column (2) uses an OLS regression to estimate the effect on the untransformed number of working days. Column (3) uses a negative binomial regression. Column (4) uses the mean of processing times for applications *in that office* that were processed after the number of days that the application I am imputing the processing time for has been pending. Column (5) 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 observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.5.

Table A13: Robustness to Alternative Measures of Baseline Performance

	Time index			Typical payments			Reported payments		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Scorecard × Overperform 3m baseline	0.017***			1,882***			595**		
	(0.080)			(710)			(230)		
Scorecard × Underperform 3m baseline	0.248***			-9.1			38		
	(0.087)			(946)			(258)		
Treat x 75-100th percentile		-0.027			1,696*			604**	
		(0.100)			(963)			(265)	
Treat x 50-75th percentile		0.044			2,435**			669*	
		(0.117)			(1,163)			(383)	
Treat x 25-50th percentile		0.135			116			242	
		(0.119)			(1,048)			(312)	
Treat x 0-25th percentile		0.337***			-143			-116	
		(0.121)			(1,496)			(391)	
Treat x Baseline ranking			-0.007**			38			15*
			(0.003)			(29)			(8.6)
Scorecard			0.115**			958			330*
			(0.057)			(609)			(173)
P-value subgroup diff.	0.05			0.11			0.11		
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	1,034,688	1,034,688	1,034,688	1,802	1,802	1,802	2,869	2,869	2,869
Clusters	311	311	311	112	112	112	112	112	112

Notes: 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; a higher index value indicates a faster processing time. Columns (4)-(6) show the effect on the estimate for how much is it "normal for a person like yourself to pay" 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 land 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 observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.5.

Table A14: Treatment Effects on Expected Processing Time

	ln(Expected processing time)				
	(1)	(2)	(3)	(4)	(5)
Scorecard	-0.006		-0.000		
	(0.035)		(0.041)		
Information treatment		-0.049**	-0.044*		
		(0.022)	(0.024)		
Scorecard x Information			-0.009		
			(0.042)		
Scorecard x Overperform				0.046	
				(0.050)	
Scorecard x Underperform				-0.041	
				(0.049)	
Info. x Scorecard x Overperform					0.018
					(0.056)
No info. x Scorecard x Overperform					0.073
					(0.057)
Info. x Scorecard x Underperform					-0.067
					(0.051)
No info. x Scorecard x Underperform					-0.014
					(0.057)
Overperform baseline				-0.129**	-0.129**
				(0.058)	(0.058)
Start-month FE	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes
Observations	2,467	2,467	2,467	2,467	2,467
Clusters	112	561	112	112	112
Control mean	61	68	62		

Notes: This table reports the effect of the scorecard and information treatments on expected processing times at the time of the in-person interview. The outcomes variable is the log transformation of the sum of the expected future processing time and the processing time already incurred at the time of the in-person interview, winsorized at the 99th percentile. Standard errors are clustered at the land-office level, except for in column (2), where standard errors are clustered at the land-office-by-survey-day level, as that is the level of randomization for the information treatment. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 6.3.

Table A15: Scorecards' Effect on Bureaucrat Transfers

	(1) Transfer	(2) Transfer	(3) Duration	(4) Duration	(5) No ACL	(6) No ACL
Scorecard	0.003		0.641		0.002	
	(0.006)		(0.707)		(0.023)	
Scorecard × Overperform		0.007		0.393		0.008
		(0.008)		(0.970)		(0.032)
Scorecard × Underperform		-0.002		0.928		-0.006
		(0.008)		(1.057)		(0.035)
Overperform baseline		-0.001		-0.285		0.013
		(0.010)		(1.182)		(0.033)
P-value: subgroup diff.		0.46		0.71		0.78
Month FE	Yes	Yes	No	No	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,516	4,516	306	306	4,516	4,516
Clusters	311	311	306	306	311	311
Control mean	0.07		12.54		0.13	
Overperformers: control mean		0.07		12.60		0.12
Underperformers: control mean		0.07		12.46		0.13

Notes: This table reports the effect of the scorecards on transfers of ACLs. Columns (1) and (2) show the effects on the percentage 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. See discussion in Section C.3.2.

Table A16: Treatment Effects on Survey Attrition

	Attrition: Survey				Attrition: Typical bribe			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Scorecard	0.027*			0.029	-0.006			0.001
	(0.015)			(0.021)	(0.020)			(0.030)
Scorecard × Overperform baseline		0.034*				0.005		
		(0.017)				(0.029)		
Scorecard × Underperform baseline		0.022				-0.017		
		(0.026)				(0.030)		
Overperform baseline		-0.017				-0.014		
		(0.022)				(0.031)		
Information treatment			-0.002	-0.001			0.020	0.029
			(0.012)	(0.018)			(0.024)	(0.034)
Scorecard x Information				-0.004				-0.017
				(0.024)				(0.047)
Start-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.09		0.11	0.09	0.44		0.43	0.43
Overperformers: Control mean		0.08				0.42		
Underperformers: Control mean		0.10				0.46		
Observations	3,213	3,213	3,213	3,213	3,213	3,213	3,213	3,213
Clusters	112	112	112	112	112	112	112	112

Notes: This table reports the effect of the scorecards and information treatment on attrition from the survey. In columns (1)-(4), attrition is measured from an applicant being approached for the in-person survey until having a successful follow-up survey by phone. In columns (5)-(6), attrition is measured from an applicant being approached for the in-person survey until the respondent answered the question about the typical bribe payment. Standard errors are clustered at the land-office level. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section B.3.4.

Table A17: Lower Lee Bounds for Scorecards' Effects on Bribes

	Typical payment		Reported payment	
	(1)	(2)	(3)	(4)
Scorecard	871		233	
	(617)		(178)	
Scorecard × Overperform baseline		1,627**		546**
		(729)		(219)
Scorecard × Underperform baseline		-105		-23
		(967)		(255)
Overperform baseline		-1,830*		-808***
		(969)		(292)
Start-month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	1,808	1,829	2,922	2,922
Clusters	112	112	112	112
Control mean	6,124		1,284	
Overperformers: Control mean		5,448		944
Underperformers: Control mean		6,726		1,578

Notes: This table reports the lower Lee bounds (Lee, 2009) for the estimates of the effects of the scorecards on bribe payments shown in columns (1) and (2) of Table 3 and column (2) and (3) of Table 4. If the estimated effect of the scorecards on attrition in Appendix Table A16 was positive, a number of randomly selected observations equal to the differential attrition rate are added back into the treatment group data and are assigned a value of zero for bribe payments. If the estimated effect of the scorecards on attrition was negative, a number of randomly selected observations equal to the differential attrition rate, are added back into the control group data and are assigned the maximum reported bribe as their bribe payments. The outcome variables are winsorized at the 99th percentile. Standard errors are clustered at the land-office level. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Appendix B.3.6.

Table A18: Comparison of Effects in Administrative and Survey Data

	≤45 working days		ln(working days)	
Data source:	Admin.	Survey	Admin.	Survey
Panel A: Overall effects	(1)	(2)	(3)	(4)
Scorecard	0.055	0.047	-0.094	-0.062
	(0.060)	(0.034)	(0.131)	(0.053)
Observations	119,706	2,260	119,706	2,260
Clusters	111	111	111	111
Control mean	0.47	0.56	85.09	59.80
Panel B: Heterogeneous effects				
Scorecard × Overperform	-0.038	0.006	0.018	0.007
	(0.081)	(0.054)	(0.181)	(0.085)
Scorecard × Underperform	0.126	0.073	-0.159	-0.106
	(0.090)	(0.046)	(0.203)	(0.069)
Overperform baseline	0.328***	0.150***	-0.519**	-0.248***
	(0.103)	(0.053)	(0.225)	(0.081)
Start-month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	119,706	2,260	119,706	2,260
Clusters	111	111	111	111

Notes: This table compares the effects estimated using the administrative and survey data. The data includes applications made in the same land offices and during the same period (February 2, 2019 to May 26, 2019). The regression specifications are the same as in Tables 2 and 4. Columns (1) and (3) show the results estimated using the administrative data. Columns (2) and (4) show the results estimated using the survey data. One land office is dropped from the analysis because no applications were made in the e-governance system in this office during the survey period. Standard errors are clustered at the land-office level. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Appendix B.3.7.

Table A19: Scorecards' Effect on Processing Times for Applications Not Affected by Survey

Panel A: Overall effect	(1)	(2)	(3)
	$\leq$45 w. days	ln(w. days)	Time index
Scorecard	0.068**	-0.172**	0.140**
	(0.029)	(0.067)	(0.059)
Panel B: Heterogeneous effects			
Scorecard $\times$ Overperform	0.015	-0.043	0.031
	(0.041)	(0.088)	(0.082)
Scorecard $\times$ Underperform	0.126***	-0.313***	0.260***
	(0.040)	(0.098)	(0.082)
Overperform baseline	0.234***	-0.531***	0.478***
	(0.055)	(0.136)	(0.112)
Start-month FE	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes
Observations	541,681	541,681	541,681
Clusters	310	310	310
Control mean	0.52	78.33	0.00

Notes: This table reports the results from Table 2 and column (1) of Table 4 when restricting the sample to applications that were made either in offices where the applicant survey did not take place, or made one month or more before the start of the applicant survey. Hence, these results are 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 observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Appendix C.1.

Table A20: Scorecards' Effects on Applications Received and Land Size

	ln(Applications received)	ln(Land size)
Panel A: Overall effects	(1)	(2)
Scorecard	-0.053	-0.028
	(0.073)	(0.063)
Panel B: Heterogeneous effects		
Scorecard $\times$ Overperform	-0.062	0.025
	(0.107)	(0.092)
Scorecard $\times$ Underperform	-0.044	-0.087
	(0.106)	(0.091)
Overperform baseline	0.147	-0.088
	(0.123)	(0.110)
Start-month FE	No	Yes
Stratum FE	Yes	Yes
Weighted by office	No	Yes
Observations	311	1,007,014
Clusters		311
Control mean	3,153	0

Notes: This table reports the effect of the scorecards on the number of applications received and the land size of those applications. In column (1), observations are at the land-office level. In column (2), observations are at the application level. Panel A shows the estimates of the overall effect and Panel B shows the estimates of the heterogeneous effects. Data contains all applications made between one month before the experiment started and 45 working days before the experiment ended (August 13, 2018 to January 20, 2020). Standard errors are clustered at the land-office level. Observations in column (2) are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Appendix C.2.

Table A21: Scorecards' Effects on Rejection Rates

Panel A: Overall effect	(1) Rejection	(2) Previously rejected
Scorecard	-0.000	0.024
	(0.021)	(0.020)
Panel B: Heterogeneous effects		
Scorecard × Overperform baseline	0.014	0.064
	(0.029)	(0.040)
Scorecard × Underperform baseline	-0.016	-0.020
	(0.031)	(0.013)
Overperform baseline	-0.068*	-0.004
	(0.038)	(0.029)
P-value subgroup diff.	0.49	0.06
Start-month FE	Yes	Yes
Stratum FE	Yes	Yes
Weighted by office	Yes	Yes
Observations	1,034,688	3,064
Clusters	311	112
Control mean	0.32	0.06
Overperformers: Control mean	0.28	0.06
Underperformers: Control mean	0.36	0.05

Notes: This table reports the effect of the scorecards on rejections of applications for land record changes. Column (1) show the effect of the scorecards on the percentage of applications rejected in the administrative data. Column (2) show the effect on the percentage of applicants surveyed who were returning after having had their application rejected, which is a proxy for incorrect rejections. Panel A shows the estimates of the overall effect and Panel B shows the estimates of the heterogeneous effects. 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. See discussion in Appendix C.2.2.

Table A22: Spillover Effects on Applications Made Outside of E-Governance System

Panel A: Overall effect	(1) Visits	(2) $\leq$45 w. days	(3) ln(w. days)	(4) ICW Index
Scorecard	-1.11	0.00	-0.13	0.05
	(0.79)	(0.06)	(0.10)	(0.11)
Panel B: Heterogeneous effects				
Scorecard $\times$ Overperform baseline	1.10	-0.07	-0.00	-0.13
	(1.59)	(0.12)	(0.22)	(0.19)
Scorecard $\times$ Underperform baseline	-1.47	0.07	-0.19**	0.18
	(1.38)	(0.06)	(0.09)	(0.14)
Overperform baseline	-2.13	0.02	-0.23	0.18
	(2.25)	(0.12)	(0.22)	(0.22)
Start-month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Observations	644	644	644	644
Clusters	58	58	58	58
Control mean	10.53	0.45	75.30	

Notes: This table reports the spillover effects of the scorecards among applications made outside the e-governance system on visits to land offices and processing times, as measured in the survey data. Only 58 of the 112 land offices had at least one application made outside the e-governance system in the survey. Standard errors are clustered at the land-office level. Observations are inversely weighted by the number of observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section C.2.3.

Table A23: Robustness of Information Treatment Effect to Different Treatment Variables

Panel A:	Typical bribe payment			
	(1)	(2)	(3)	(4)
Information treatment	212	120	211	221
	(490)	(493)	(501)	(548)
Observations	1,802	1,733	1,802	1,448
Clusters	539	536	539	437
Panel B:	Typical payment in scorecard offices			
Information treatment	300	501	300	498
	(775)	(802)	(775)	(865)
Observations	930	881	930	785
Clusters	271	269	271	234
Panel C:	ln(Expected processing time)			
Information treatment	-0.049**	-0.044*	-0.007	-0.029
	(0.022)	(0.022)	(0.022)	(0.024)
Observations	2,467	2,467	2,467	2,064
Clusters	561	561	561	462
Start-month FE	Yes	Yes	Yes	Yes
Stratum FE	Yes	Yes	Yes	Yes
Weighted by office	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Correct

Notes: This table shows the robustness of the estimates of the effects of the information intervention to defining the treatment variable differently with respect to noncompliance with the treatment assignment in the treatment delivery. The results investigated is the overall effect on bribe payments, the effects on bribe payments among offices receiving the scorecard treatment, and the effect on the expected processing time at the time of the in-person survey. Column (1) shows the results when using my preferred treatment variable based on the median treatment delivered in a land office survey day. Column (2) shows the results when using the actual treatment delivered for each applicant. Column (3) shows the results when using the assigned treatment for each land office survey day. Column (4) shows the results when using the assigned treatment but restrict the sample to applicants who received the assigned treatment. Outcome variables are winsorized at the 99th percentile. Standard errors are clustered at the land-office level. ***p<0.01; **p<0.05; *p<0.1. See discussion in Appendix C.4.1.

Table A24: Scorecards' Effect on Applicant Satisfaction

	(1) Satisfaction	(2) Satisfaction
Scorecard	-0.038	
	(0.060)	
Scorecard x Overperform		-0.080
		(0.081)
Scorecard x Underperform		-0.016
		(0.086)
Overperform baseline		0.181*
		(0.097)
P-value subgroup diff.		0.60
Start-month FE	Yes	Yes
Stratum FE	Yes	Yes
Weighted by office	Yes	Yes
Observations	2,869	2,869
Clusters	112	112

Notes: This table reports 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 observations in that land office. ***p<0.01; **p<0.05; *p<0.1. See discussion in Appendix C.6.