Public Sector Salaries and the Quality of Governance: Evidence from Frontline Bureaucrats in India

Siddharth Eapen George & Martin Mattsson*

December 18, 2024

Abstract

Salaries are the largest expenditure item for most governments. We examine the effects of increasing bureaucrats' pay, by analyzing a policy change in Telangana, India, where one group of frontline bureaucrats received a 91% pay increase while another group of bureaucrats performing the same job did not. Using a difference-in-differences design, we show that, on average, higher salaries (i) had no impact on performance, (ii) did not affect dishonest or corrupt behavior, (iii) reduced quit rates by 2.4 percentage points (16%) after 2 years but did not affect the average quality of bureaucrats in service. Higher salaries marginally increased effort among bureaucrats who were unhappy about their pay. Expert forecasters incorrectly predicted that higher salaries would improve both bureaucrat performance and selection.

JEL Codes: D73, H76, O10

Keywords: Public Sector Pay, Bureaucracy, Panchayat Secretaries

*National University of Singapore, segeorge@nus.edu.sg and martin.mattsson@nus.edu.sg. We are deeply indebted to Karthik Muralidharan and the Centre for Governance of Indian States (CEGIS) for institutional support and data access, without which this project would not have been possible. We would like to thank Kory Kroft, Biju Rao, Praveen Ravi, Yogita Shamdasani, Diego Salazar, Yusuf Neggers, and Ao Wang for their insightful comments and feedback. Andrew Tiong Jia Siang provided outstanding research assistance. We acknowledge financial support from the National University of Singapore.

1 Introduction

Public sector salaries are the largest expenditure item in most government budgets, accounting for 27.4% of public expenditure and 9.8% of GDP in the average country (World Bank, 2023). Understanding the effects of changing bureaucrats' salaries is thus a first-order public finance question.

Most bureaucrats receive fixed salaries that are not directly linked to their performance. Economic theory makes ambiguous predictions about the impacts of raising fixed salaries. On one hand, higher salaries may improve selection by enabling the public sector to attract and retain high-ability citizens (Dal Bó et al., 2013). Higher salaries may also improve the performance of incumbent bureaucrats by raising morale (Mas, 2006) and triggering efficiency wage dynamics (Becker and Stigler, 1974). On the other hand, incumbent bureaucrats may not respond to higher salaries at all (De Ree et al., 2018), if they face little risk of being fired and their promotion prospects are only weakly tied to performance (Iyer and Mani, 2012). Moreover, in hierarchical organizations like bureaucracies, salary increases also change vertical pay disparities, which may affect performance incentives and morale (Deserranno et al., 2024; Breza et al., 2018).

Identifying the effects of bureaucrat pay is empirically challenging. The ideal experiment would find two identical groups of bureaucrats performing the same job, randomly select one to receive higher salaries, and compare the subsequent performance of both groups (Dal Bó et al., 2013; De Ree et al., 2018). However, such variation is rare, since bureaucrats' salaries are typically set by rules or formulas that apply to all bureaucrats and seldom change.

In this paper, we analyze a policy change in the Indian state of Telangana, where the salaries of some but not all *Gram Panchayat* (GP) secretaries were increased. GPs are the lowest tier of government in rural India, responsible for delivering local public services and running government schemes. Each GP is governed by an elected council, but its decisions are implemented by a Panchayat Secretary (PS), who handles the GP's day-to-day operations and administration. PSes hire local staff, maintain public infrastructure (e.g. roads, streetlights), administer public services (e.g. sanitation), implement development programs, and manage public funds. India has approximately 250,000 GPs serving more than 900 million citizens, and each is administered by a PS. Despite their prevalence and importance, which has been highlighted in qualitative research (Rao et al., 2017; Verma and Ranjan, 2023) and even popular culture (e.g. this popular comedy-drama), PSes remain virtually unstudied by economists.

In Telangana, PSes were recruited in several waves. Secretaries recruited in the most recent wave (*Junior PSes*) received a different contract and lower salary compared to those recruited in earlier waves (*Regular PSes*, a group comprising secretaries in four higher salary grades). However, all secretaries performed exactly the same duties in their respective GPs.[1] In July 2021, the Telangana government raised the monthly salaries of Junior PSes from INR 15,000 (USD 195) to INR 28,719 (USD 374), a 91% increase. By contrast, the salaries of Regular PSes, which ranged from Rs 28,719 (Grade IV PSes) to Rs 42,300 (Grade I PSes), were left unchanged. Moreover, the policy change was sudden—it was announced on 16 July 2021 and took retroactive effect from 1 July 2021. We evaluate the impact of the salary increase using a difference-in-differences (DiD) design, which compares (i) Junior PSes and Regular PSes, (ii) before and after the pay increase.

We examine the impacts of higher salaries on three categories of outcomes: (i) *performance*, (ii) *honesty* and *corruption*, and (iii) *selection*. Measuring bureaucrat performance is challenging, especially for core civil service jobs (like GP secretary) that have multiple responsibilities. We worked with Telangana's Department of Panchayat Raj and Rural Development (PRRD) to construct performance measures that reflected GP secretaries' deliverables, contained outcomes they could influence, and were hard to manipulate. Specifically, we leverage evaluations of secretaries by (i) supervisors, (ii) third-party auditors, and (iii) citizens. These different evaluations are all positively correlated, suggesting that they capture some core elements of secretary performance. We also construct a proxy measure of bureaucrat effort, by measuring how regularly secretaries submit a mandatory daily report. Collectively, these data sources provide a holistic, 360-degree view of a secre-

[1]Junior PSes comprise 71% of secretaries in Telangana, while Regular PSes comprise the remaining 29%. There is another salient difference between Junior PSes and Regular PSes contracts: JPSes were hired on 4-year contracts, after which they could be made tenured public servants conditional on satisfactory performance. By contrast, Regular PSes were hired directly into tenured positions.

tary's performance. These outcomes show parallel trends before the reform, supporting the validity of the DiD design.

We find that higher salaries had no impact on bureaucrat performance, whether measured by supervisor, auditor, or citizen evaluations. The point estimates imply a -0.008 standard deviation (SD) change in supervisor evaluations, a 0.01 SD change in auditor scores, and a 0.02 SD change in citizen assessments. Moreover, this zero effect is precisely estimated: with 95% confidence, we can reject improvements greater than 0.023 SD for supervisor evaluations — the department's official metric for assessing secretaries.[2] Turning to effort, we find that higher salaries led to a 0.71 percentage point decrease in the daily reporting rate. While statistically significant, this effect (about 0.03 SD or 1%) is too small to be economically meaningful.

Second, we study how the salary increase affected dishonesty and corruption. Our measure of honesty captures the extent to which secretaries misreport data to inflate their performance. We detect misreporting by comparing self-reported data with supervisor assessments of the same indicators. At baseline, secretaries inflate their performance by an average of 19%. However, we find that higher salaries had no impact on inflation. We also construct two measures of corruption. First, we develop an indicator of nepotism in GP hiring — the share of hired GP staff with the same name as the GP's political leader. A similar measure is used by Lehne et al. (2018) to proxy for corruption in contract allocation in the Indian context. Second, we measure the share of GP expenditure spent on reimbursing the secretary for personal expenses. For both corruption measures, we find that higher salaries have a precisely estimated zero effect.

Third, we study how higher salaries affect the selection of individuals working in the bureaucracy. Using administrative data on the roster of PRRD bureaucrats, we track which PSes are still in service two years after the pay increase. Our DiD estimates suggest that higher salaries decreased the probability of leaving the bureaucracy by about 1 percentage point per year. Approximately 14.9% of incumbent GP secretaries in April 2020 had left the PRRD Department by March 2023. We estimate that this share would have been

[2] We are only able to rule out effects larger than 0.100 SD for auditor assessments and 0.165 SD for citizen evaluations as we have a significantly smaller sample for auditor and citizen evaluations.

2.4 percentage points (15.9%) higher had salaries not been raised. However, we find no evidence that higher salaries changed the composition of bureaucrats who remained in service. Before the policy change, performance was uncorrelated with quitting the service, and this relationship does not change after the salary hike. Treatment effects on quit rates also do not vary with most secretary characteristics (e.g. baseline performance and caste). An exception is gender: higher salaries increase stay rates of men and have no impact on those of women. Thus, pay increases did not change the quality of the bureaucracy in the short run and had a limited impact on its composition, except for exacerbating the over-representation of men.

We asked 85 academics and policy professionals on the Social Science Prediction platform to forecast our effects. Most predicted that higher salaries would have positive effects. On average, respondents predicted that the salary increase for GP secretaries would (i) improve performance (as measured by supervisor assessments) by 0.21 SDs, (ii) reduce self-dealing, (iii) decrease quit rates, and (iv) improve the average quality of the bureaucrats who stayed in the job. We only find empirical support for prediction (iii). Forecasts are similarly inaccurate when we restrict the sample to respondents who are familiar with the academic literature on public sector pay, those familiar with the Indian context, or those with a PhD.

We explore two mechanisms through which pay increases may affect bureaucrat effort and performance: morale and incentives. Pay hikes could raise morale, but may also weaken performance incentives by narrowing the gap between a bureaucrat's salary and her superior's. Our reduced form results identify the net impact of both forces, and suggest that, on average, they are either small or cancel out in our context. Nevertheless, morale or incentive effects may have been particularly strong for some bureaucrats.

The scale of the policy change, which affected 12,770 secretaries in GPs with a combined population of over 21 million, gives us power to test for treatment effect heterogeneity. We find that bureaucrats who were dissatisfied at baseline, particularly the 17.4% of PSes who had complained about their salaries in a pre-reform feedback exercise, responded more positively to the pay increase. Higher salaries induced additional effort from these officers. By contrast, Junior PSes who were concerned about being promoted to match the pay and terms of Regular PSes responded negatively to the pay increase. Higher salaries reduced their effort, possibly by dampening the incentive to perform well and get promoted. Neither traditional nor machine learning methods identify any demographic characteristics (e.g. gender, caste) as robust correlates of treatment effect heterogeneity.

Having shown that raising bureaucrats' absolute salaries has little effect on performance, we investigate the impact of changing relative pay, which prior research links to worker morale and incentives (Card et al., 2012; Breza et al., 2018; Deserranno et al., 2024). To isolate the effects of pay rank, we focus on a group of bureaucrats whose absolute pay was unaffected by the reform, but whose rank in the pay hierarchy was. Grade IV PSes occupied the position in the salary scale just above Junior PSes. The pay hike for Junior PSes equalized their salaries, placing both at the bottom of the salary pyramid. By contrast, the reform did not affect the pay rank of Grade I-III PSes, who were paid significantly more in higher salary grades. Using a difference-in-differences design, we compare the performance of (i) Grade IV PSes and Grade I-III PSes, (ii) before and after the Junior PS pay increase. We find no evidence that the decrease in relative pay affected the performance or effort of Grade IV PSes, suggesting that changes in pay rank had limited impact in our context.

Contribution to literature.

Our paper contributes to the literature on the personnel economics of the state (Finan et al., 2017) in two ways. First, we add to the literature on corruption. Since Becker and Stigler (1974), economists have proposed raising bureaucrat salaries to reduce corruption. Yet we have limited evidence on how pay affects corruption and other forms of dishonest behavior. Besides inconclusive cross-country correlations (Rauch and Evans, 2000; Van Rijckeghem and Weder, 2001; Demirgüç-Kunt et al., 2023), there is limited evidence on this question. Di Tella and Schargrodsky (2003) finds that better-paid procurement officers are less corrupt during intermediate levels of audit intensity, but their analysis relies on potentially endogenous cross-sectional wage variation, since bureaucrats' salary levels do not change during the study period.[3] By contrast, we use exogenous variation in salaries to

[3] Using a DiD design, Foltz and Opoku-Agyemang (2015) find that higher salaries for police officers increase corruption in Ghana. However, they advise readers to interpret their results cautiously due to non-parallel pre-trends before the salary change.

show that pay had no impact on either data misreporting or two measures of self-dealing. Our findings also differ from those of Niehaus and Sukhtankar (2013), who show that the opportunity to earn higher illicit returns in future reduces corruption today.

Second, and most directly, we shed light on the impacts of public sector pay. Several influential papers explore how salaries affect bureaucrat selection or on-job performance, presenting mixed results.[4] On the selection margin, Dal Bó et al. (2013) shows that higher salaries attract more qualified candidates, while Willén (2021) finds no impact on retention or recruitment. Examining incumbent behavior, de Ree et al. (2017) shows that higher base salaries had no impact on the performance of incumbent teachers, whereas Mas (2006) finds that smaller-than-expected pay increases worsened police performance.[5]

We make four contributions to this literature. First, we analyze a wider set of outcomes than prior studies, including supervisor, citizen and auditor evaluations, proxies for dishonesty and corruption, and a direct behavioral measure of effort. Unlike prior work, we study how higher salaries affect both selection and incumbent behavior in the same setting, finding increased retention but no change in selection or performance — results which differ markedly from expert forecasts. While we cannot assess how higher salaries changed entry into the public service, we believe improved retention is an important effect in its own right, as turnover may hinder public service delivery (Akhtari et al., 2022).

Second, we show that bureaucrats respond differently to pay increases: while higher salaries have no average effect on effort, they positively impact those dissatisfied with their pay. This helps reconcile the conflicting findings in de Ree et al. (2017) and Mas (2006).

Third, we consider how salary changes may have different effects in hierarchical organizations like a bureaucracy. We find suggestive evidence that higher salaries dampen incentives for lower-level bureaucrats, echoing the findings of Deserranno et al. (2024). However, changes in relative pay do not appear to affect the behavior of upper-level bureaucrats, consistent with Card et al. (2012).

Finally, we study bureaucrats in "core civil service" positions. These jobs have broad

[4] Best et al. (2023) and Barteska (2024) both show the importance of high quality individual bureaucrats in policy implementation.

[5] In addition, Cabrera and Webbink (2020) find that higher salaries for teachers in disadvantaged areas lead applicants to sort to those locations; Jayaraman et al. (2016) report that higher government-stipulated baseline salaries raised the productivity of tea pickers in the short-term but had no long-term effects.

work scopes and multi-faceted responsibilities, which make it difficult to incentivize or even measure performance (Holmstrom and Milgrom, 1991; Baicker and Jacobson, 2007; Hasnain et al., 2014; Mattsson, 2024). Given these challenges, research has tended to focus on government employees with clear tasks and easy-to-quantify performance metrics, such as teachers (Muralidharan and Sundararaman, 2011; Duflo et al., 2012; De Ree et al., 2018), frontline health workers (Ashraf et al., 2020; Hanna and Wang, 2013; Khan, 2023; Deserranno et al., 2024), and tax collectors (Khan et al., 2016, 2019). Few studies evaluate how personnel policies impact bureaucrats in core civil service positions (Rasul and Rogger, 2018; Bertrand et al., 2020), even though these jobs comprise 34% of public employment globally and are often the most important jobs in the public sector (World Bank, 2023).[6]

The remainder of the paper is structured as follows: Section 2 explains our context, Section 3.4 describes our data and empirical strategy, Section 4 presents our results, Section 5 discusses heterogeneity, and Section 6 concludes.

2 Context and Data

2.1 Local Government in Rural India

Our natural experiment takes place in Telangana, an Indian state with a population of 35 million, 61% of whom live in rural areas. The state is divided into 33 districts. Each district comprises on average 16 mandals, and each mandal comprises about 24 Gram Panchayats (GPs). The GP is the smallest unit of administration in the system of local government in rural India. GPs are responsible for a variety of local governance functions, including delivering public services and implementing development programs (George et al., 2024). Each GP is governed by a council of elected representatives led by the *Sarpanch* (or President). However, day-to-day management and administration is handled by a GP Secretary, a bureaucrat who functions as a chief operating officer and leads a small team of 2-4 staff.

[6]In comparison, teachers and medical workers together represent 11% of public employees in India, and 13% globally (World Bank, 2023).

2.2 The Role of a GP Secretary

In 2018, the Government of Telangana enacted a "Panchayat Raj" Act to codify the responsibilities of GPs and the role of a GP secretary. The Act tasked secretaries with a range of responsibilities, including overseeing public works, maintaining public facilities (e.g. roads and drains), delivering public services (e.g. sanitation and garbage collection), and handling administrative tasks such as collecting property taxes and issuing certificates.

Given the wide range of tasks, the act also defined Key Result Areas (KRAs) that highlighted GP secretaries' most important responsibilities. Performance metrics were identified for each KRA and communicated to secretaries. Appendix Table A3 describes each KRA and performance metric. To give an illusrative example, streetlight maintenance was one KRA and the GP secretary's performance metric was to ensure at least 90% of streetlights were working in their GP.

Like most bureaucrats, the PS reports both to an elected official (the Sarpanch) and a bureaucratic supervisor — the Mandal Panchayat Officer, who oversees all ~20 secretaries and the delivery of public services in their mandal. Each supervisor inspects each GP under their charge at least once per quarter, and writes a report evaluating the performance of the GP secretary. Auditors (called State Level Visitors) also visit GPs to inspect the quality of public infrastructure and service delivery.

2.3 Policy Change: Salary Increase for GP Secretaries

The Act mandated that every GP had a secretary. To implement this, the state appointed 9,355 new secretaries in a recruitment drive in 2018-2019. These newly-hired Junior Panchayat Secretaries (PSes) were appointed on 4-year contracts, and were to be regularized as permanent government employees conditional on good performance. By April 2021, Telangana had 12,101 secretaries, with Junior PSes handling 9,081 GPs and Regular PSes overseeing 2,839 GPs. Regular PSes were hired in prior recruitment waves, with starting dates ranging from 1985 to 2018.

Junior PSes and Regular PSes did the identical job in different GPs, with the same responsibilities, powers, and performance metrics. However, the Junior PSes were initially paid INR 15,000 per month (USD 195) while the Regular PSes received monthly salaries between INR 28,719 (USD 374) and INR 42,300 (USD 503), depending on their salary grade.[7] Each regular PS was in one of four salary grades — the lowest being Panchayat Secretary Grade IV and the highest being Panchayat Secretary Grade I — roughly, but not completely, based on their seniority.[8] Pay-related concerns were commonly raised by secretaries when providing feedback to the department. In an open-ended question about issues faced in their jobs, 17.4% mentioned insufficient pay. Moreover, Junior PSes were twice as likely as Regular PSes to mention salary-related concerns (20% compared to 9%), reflecting their substantially lower pay. Further details about the survey are in Appendix A.3.

On 15 July 2021, the government announced that it would raise the pay of Junior PSes from INR 15,000 (USD 195) to INR 28,719 (USD 374), a 91% increase. The pay change raised the salaries of Junior PSes to be equal to those of the lowest paid regular PSes (Grade IV PSes). The salary rise was unexpected and took effect immediately, with the new higher salaries taking retroactive effect from 1 July 2021. By contrast, salaries for regular PSes of all grades were left unchanged. We exploit this policy change to examine the impact of higher salaries on GP secretary performance and selection.

While the salary increased by 91%, some Junior PSes may have already expected that there was a chance their salary would eventually increase if they were regularized (i.e. offered permanent contracts) at the end of their 4-year contracts. In Appendix A.4, we calculate approximately how much the expected net present value of bureaucrats' incomes changed as a result of the salary increase. Using reasonable assumptions about Junior PSes' expectations, we find that the pay increase raised the net present value of bureaucrats' salaries by between 30% to 40%.

When the salary for the Junior PSes increase, the salary scale becomes more compressed and the incentives for the Junior PSes to get regularized decrease. This is a typical feature of any salary change in a hierarchical organization with opportunities for promotions. Unless the whole salary scale is changed in a way that preserve the difference in utility

[7] Throughout the paper we use USD/INR=76.8, the average daily exchange rate from January 2021 to Mar 2023.

[8] A small number of GPs were administered by Junior PSes hired on an outsourcing basis. These PSes did not receive a pay increase and thus we do not classify them as Junior PSes. Removing these GPs from the sample does not meaningfully change our results.

between the different levels in the organization, the salary change will also affect the promotion incentives. Therefore, the pay policy charge we are evaluating is a typical change considered by governments.

2.4 Baseline Differences Between Regular and Junior GP Secretaries

Appendix Table A1 shows baseline summary statistics for the Junior PSes (who were eligible for the pay increase) and Regular PSes (who were not), while Appendix Table A2 details characteristics of the GPs they serve.[9] Junior PSes were more likely to be in smaller GPs with marginally lower levels of public good provision. Demographically, Junior PSes are more likely to be female and from disadvantaged social groups (Scheduled Castes and Tribes). Moreover, since they were recruited more recently, Junior PSes are unsurprisingly younger and have fewer years of service in the PRRD. They also have slightly better baseline performance scores on some dimensions, but were more likely to be dissatisfied and to have complained about their salaries. However, these differences do not pose a problem for our identification strategy, as we do not assume that the salary increase was randomly allocated to PSes. Instead, as described in Section 3.4, our identifying assumption is that GPs with Junior and Regular PSes were on common trends in the absence of the salary increase.

3 Data and Empirical Strategy

Measuring the performance of GP Secretaries is challenging. Their work is multi-dimensional and contains tasks that are hard to objectively verify. These are common challenges when measuring the performance of almost any bureaucrat above the lowest level of frontline providers (Besley et al., 2022). We overcome these challenges by using rich data from multiple stakeholders, containing both objective and subjective performance assessments. In collaboration with the PRRD Department, we designed a performance measurement system to capture the key elements of secretary performance and facilitate the measurement

[9]The village characteristics used in this table come from SHRUG (Census of India, 2011; Socio Economic and Caste Census, 2011; Henderson et al., 2012; Asher et al., 2021; Chi et al., 2022).

of data misreporting.[10]

3.1 Data: Performance Indicators

We construct four main measures of secretary performance. The measures evaluate the secretary from different perspectives and complement each other to give a holistic assessment. Appendix Figure A1 shows the pairwise correlation between the key performance measures. Almost all pairwise correlations are positive and statistically significant. This suggests that the different performance measures can be thought of as different signals of some true underlying performance.

Supervisor Evaluation. The first and arguably most important measure is based on assessments by the secretaries' direct supervisor in the bureaucracy. The supervisor is required to visit each GP at least once a quarter (in practice GPs are visited once every 3.6 months on average). The supervisor assesses the quality of public services in the GP, focusing on the KRAs highlighted in the Act. The supervisor evaluates (i) the cleanliness of roads, drains, water tanks, and institutions like primary schools, (ii) whether or not composting procedures have been followed, (iii) whether and what share of streetlights are working, (iv) whether plantation nurseries have sufficient infrastructure and fencing, (v) plant survival in nurseries and (vi) the quality of GP administration, including maintenance of registers such as birth and death records and organization of GP-wide citizen assemblies. We aggregate these indicators in the same way as the state government aggregates them and convert this into a performance score ranging from 0 to 100 percent. These scores for the basis of the department's official evaluation of GP secretaries.[11] See Appendix A.2 for a more detailed description of how the data is processed. In addition to these objective performance indicators the supervisor also answers a question about how well the secretary is performing overall, this is what we call the *subjective performance assessment*. Our results are consistent across these two performance measures.

[10]This data was also used to experimentally evaluate the impact of a performance scorecard. The pre-analysis plan of the experiment can be found via the AEA Registry here.

[11]When not all indicators are available for a given inspection, we calculate the percentage score by dividing the sum of the available indicators by the total potential score, which is the sum of the maximum values of the available indicators.

Auditor Assessment. Our next performance indicator is based on evaluations by state-level auditors. The state-level auditor assesses the GP along similar dimensions as the supervisor, though they omit certain categories (for example, auditors do not have the jurisdiction to examine GP records and assess administration). There are substantially fewer auditors than supervisors, and they audit staff other than GP secretaries as well. Thus, during our sample period (Jan 2021 - Dec 2022), a total of 9,994 GPs were audited (78% of all GPs), and the average GP was visited by an auditor once every 8.1 months.[12] As with supervisor evaluations, we construct an auditor performance score from 0 to 100 percent based on the weights on different KRAs given by the PRRD Department.

Citizen Assessment. The third performance indicator is a citizen survey. 10,606 citizens were surveyed in a total of 4,672 GPs (about 37% of all GPs) over 4 months, covering a period before and after the pay change (May-Aug 2021). These surveys took place via phone and were conducted by a survey company contracted by the PRRD Department. GPs were first randomly sampled from all mandals. Then, within each sampled GP, citizens were randomly sampled based on phone numbers in the databases for property records and beneficiaries for subsidized food. These databases together cover over 90% of rural households.

Effort. Our final performance measure is a behavioral measure of GP secretary effort. All secretaries are required to submit a daily report via a mobile app. Each day, the app prompts the secretary to visit a road in their GP and make observations regarding (i) the cleanliness of the road, (ii) the cleanliness of the drains alongside the road, (iii) the number of streetlights on the road and the number that are working. Since this data is self-reported, we do not use the reports as objective performance outcomes themselves. Instead, we develop a measure of effort based on how often they file the required daily report.

3.2 Data: Retention and Selection

An important reason to increase public sector wages is to attract and retain high-quality workers. Our context does not allow us to study the entry margin, as the salary change

[12] However, the effective sample size is only 4,851 GPs since the remainder are singletons, visited only once by the auditor.

took place after the recruitment drive and there have been no significant recruitments since then. However, we are able to examine retention, using administrative data from four complete personnel rosters of all Panchayat and Mandal public servants from May 2020, May 2021, December 2021, and March 2023. We measure retention as a "stay rate" from the base period of May 2020, the start of our administrative data. We classify any secretary found in the roster as still in public service, regardless of their current position. We classify any individual no longer in the roster, as having left the public service.[13] The overall stay rate from May 2020 to March 2023 was 85.1%, implying a relatively low annual attrition rate of less than 5%. See Appendix A.1 for more details.

3.3 Data: Honesty and Corruption

3.3.1 Misreporting

Self-assessment. Like many frontline bureaucrats, GP secretaries do not just deliver public services, but also collect the administrative data that is used by senior officials to monitor them and understand the quality of program implementation. Besides the daily sanitation report described above, secretaries also file a report at the end of every month. The monthly statement covers a wider range of KRAs. Secretaries report on the cleanliness of water tanks and the quality of drinking water, the regularity of garbage collection and composting, and the management of the GP's plantations, nurseries and crematoriums, etc. Details on all the indicators are in Appendix Table A3. We use the department's weights for each indicator to construct a *self-assessment score* on a scale from 0 to 100.

Inflation. Since most indicators in the self-assessment score are also covered by the supervisors during their inspections, we can compare the self-reported data against the supervisor evaluations of the same indicators. In fact, the app used by supervisors to fill in the data was programmed to facilitate such back-checking. When a supervisor visits a GP, she is prompted by the app to visit the two roads most recently reported on by the secretary. In most GPs, given the reasonably high daily reporting rate (around 67%), these roads would have been visited and reported on by the secretary in the past 2-3 days.

[13] If an individual is missing from one roster and then reappears in a later roster we classify them as not having left the service.

The supervisor is then asked to evaluate the cleanliness of the road and drains alongside it, and the functioning of streetlights on the road — the same questions answered by the secretary in the preceding days. For all overlapping indicators — where we have both self-reported data and supervisor assessments — we construct a measure of inflation based on the discrepancy between the two. We define inflation as the difference between the self-reported score and the supervisor-assessed score on the overlapping indicators. To make the inflation metric more interpretable, we convert it into percentage terms by dividing it by the maximum possible score from the overlapping indicators.

$$Inflation = \frac{SelfAssessedScore - SupervisorScore}{MaxSupervisorScore}$$

The inflation variable is calculated at the GP-by-month level for the GP-months that we have both a supervisor and self-reported score for. The average inflation in our sample is 19.0%.

3.3.2 Corruption

Due to its illicit nature, corruption is notoriously hard to measure (Olken and Pande, 2012). We use two proxies to measure corruption. The first is a proxy for the GP secretary's own corruption. We use the share of the GP's budget that is being spent on reimbursing the secretary for expenses. Generating false expenses and reimbursing oneself for these is a common type of corruption and our assumption is that a GP secretary engaging in such behavior would have a higher share of the total GP expenses being spent on such reimbursements.

Second, we use the share of multipurpose workers hired by the GP that share a name with the GP's Sarpanch. A similar measure is used in the Indian context by Lehne et al. (2018) to measure corruption in the allocation of local road contracts. Hiring relatives as multipurpose workers, with very low expectations of any work being done, is a common form of corruption and nepotism. The selection of multipurpose workers is done by the Sarpanch and not by the secretary, but the secretary must process and administer the hire and disburse the salary to the worker. Thus, this measure is a proxy for the corruption that is allowed to happen by the secretaries, and not a proxy for corruption carried out directly by the secretary.

We find that in 6% of the transactions, one of the names of the multipurpose workers matches one of the names of the Sarpanch. We would expect some multipurpose workers to share a name with the Sarpanch by chance, but Appendix Figure A2 shows that when we simulate how often the names coincide by chance by matching multipurpose worker names with random Sarpanch names this can only explain approximately 1.4% of name matches.[14] Therefore, we believe that the majority of name matches are the result of some form of corruption or nepotism.

Even though neither of our proxies is a perfect measure of corruption, the fact that we are using a DiD design helps in that any time-invariant biases of the measure are differenced out.

3.4 Empirical Strategy

Our primary empirical strategy to estimate the effect of the pay increase is a difference-in-differences (DiD) design. We estimate regressions of the form:

$$y_{it} = \beta Post_t \times PayIncrease_i + \tau_t + \delta_i + \varepsilon_{it} \tag{1}$$

where y_{it} is an outcome variable for GP i in time period t. $Post_t$ indicates if time period t is after the salary increase. $PayIncrease_i$ is an indicator variable for whether the GP was run by a secretary eligible for the pay increase (i.e. a Junior PS) in May 2021, the latest pre-treatment period. For brevity, we sometimes refer to GPs with a pay-increase-eligible secretary as *Treated GPs* and GPs with a pay-increase-ineligible secretary as *Comparison GPs*. δ_i and τ_t denote GP and time fixed effects respectively. β is the coefficient of interest, estimating the average treatment effect on the treated.

Unlike some other studies on the effects of pay, we are able to estimate the effects on the quality of service even when the secretary has left the GP or the bureaucracy. This is

[14] To establish a benchmark for name sharing between multi-purpose workers and sarpanches, we first randomized sarpanch names within each Mandal. Then, we calculated the monthly average rate of name-sharing, repeating this process 100 times to mitigate random variation. For each iteration, we calculated the monthly average rate of name sharing, and these monthly averages were then aggregated across all iterations.

an advantage of our empirical strategy, since it allows us to estimate the overall (reduced form) impact of the salary increase, which may work through a combination of changing the behavior of incumbent secretaries (the incentives or morale channel) and changing the composition of secretaries (the selection channel).

The key identifying assumption is that, absent the salary increase, outcomes for GPs with a Junior PS would have on average changed similarly to those of GPs with a Regular PS. We display two graphs to test this parallel trends assumption in the period before the salary increase. First, we show timelines constructed with minimally processed data for the two groups of GPs. Second, we show the coefficients from dynamic DiD regressions of the form:

$$y_{it} = \sum_{p=\underline{T}}^{-2} 1\left[t = p\right] \times PayIncrease_i + \sum_{p=0}^{\bar{T}} 1\left[t = p\right] \times PayIncrease_i + \tau_t + \delta_i + \varepsilon_{it} \tag{2}$$

Where $1\left[t = p\right]$ are indicator variables for each time period except for the last time period before the salary change, which is kept as the comparison group. $\underline{T}$ is the first period for which we have data for outcome y, and $\bar{T}$ is the last period for which we have data. For both equations (1) and (2), we cluster standard errors at the GP level.

4 Results

We use the DiD empirical strategy to examine three main categories of outcomes — (i) performance, (ii) honesty and corruption, and (iii) retention and selection.

4.1 Performance

We examine three different measures of performance that capture the quality of public services delivered by the GP, namely evaluations by supervisors, auditors and citizens. Then, we examine how higher salaries affect bureaucrat effort. Figure 1 presents time trends of the key performance outcomes with minimally processed data. For all the measures, we observe that performance outcomes in treated and comparison GPs have similar levels and follow very similar trends prior to the salary increase, providing support for the parallel trends assumption. These trends then continue to be parallel and show no substantial deviations after the policy change, indicating that the salary increase had no impact on performance.

Supervisor Evaluations. Sub-Figure 1a presents results for supervisor assessments. There is no visible change in the gap between the lines before compared to after the salary increase, suggesting a null effect of higher salaries. Supervisor scores shows significant month-to-month variation because supervisors were learning how to use it in early 2021 and because there were occasional changes to the weights on different performance indicators (e.g. in Oct 2021). Since these factors affect both treated and comparison GPs, their impact is accounted for by the month fixed effects in Equation (1).

Appendix Sub-Figure A3a confirms the null effect of increased pay. The dynamic DiD point estimates indicate no substantial or statistically significant deviations from the parallel trends, both in the periods before and after the salary increase. Column (1) of Table 1 shows that the estimated effect of the salary increase is very close to zero, with a point estimate of -0.115 percentage points. The coefficient is also precisely estimated: with 95% confidence, we can reject a change in the supervisor assessment larger than 0.339 percentage points (0.023 SD). Moreover, we document these null effects over a period of 18 months after the pay raise, indicating that the salary increase had neither short- nor medium-term effects.

Appendix Figure A4 and Appendix Table A4 show similar results for a related outcome — supervisors' subjective overall assessment of the GP secretary, which is correlated with but distinct from the aggregated KRA-wise evaluations that form our baseline outcome.

Auditor Evaluations. Sub-Figure 1b presents time trends in auditor assessments for treated and comparison GPs. Both groups have similar pre-period levels and trends in the outcome variable, and we see no visual evidence of the salary increase impacting performance as measured by auditor evaluations.[15] Column (2) of Table 1 shows that

[15] Appendix Sub-Figure A3b confirms this using the event study specification from Equation 2. In the auditor data, June 2021 is a minor outlier, with an estimate below all the other estimates. This means that some estimates both before and after the salary increase are positive and statistically significantly different from zero when using June 2021 as the baseline month. However, there is no overall evidence of differential pre-trends or a positive effect from the salary increase.

the estimated effect of the salary increase on the auditor score is statistically insignificant and close to zero, with a point estimate of 0.171 percentage points. As there are fewer auditors and they conduct fewer inspections than supervisors, our results are somewhat less precise, but we can still reject an improvement of more than 1.280 percentage points (0.100 SD) with 95% confidence.[16]

Citizen Evaluations. Our third performance measure is based on citizen assessments of local public services and the GP secretary's performance. Sub-Figure 1c again shows parallel trends and no evidence of a clear treatment effect, though we observe outcomes over a significantly shorter time period — only 4 months as compared to over 2 years for supervisor evaluations. The dynamic DiD figure (Appendix Sub-Figure A3c) and Column (3) of Table 1 provide further support to this conclusion, that higher salaries had no impact on citizen evaluations.

Effort. Next, we assess how the salary increase impacted bureaucrat effort. Our effort proxy captures how regularly the GP secretary files the daily sanitation report, described in Section 3.3.1. Secretaries are required to submit this report daily, but the average completion rate is only 67%. Sub-Figure 1d presents the reporting rate by week for treated and comparison GPs. We see parallel trends before the policy change, and no visible change after the salary increase.

Column (4) of Table 1 shows that higher salaries led to a statistically significant decrease in effort, by 0.7pp. However, this change is small in magnitude — about 1% or 0.03 SD — which perhaps explains why it is not visually salient. In Appendix Figure A3d, we can see that when the reporting rate decreases, it tends to decrease more for the secretaries who received the salary increase than for those who did not, but in the time periods where the reporting rates are stable there are no substantial difference between the two.

4.2 Honesty and Corruption

One reason—often highlighted by economic theory—to increase salaries for public officials is to keep them honest and reduce corruption (e.g. Becker and Stigler, 1974; Di Tella and

[16]There are fewer state-level auditors than supervisors (only 110 auditors vs 540 supervisors). Moreover, the auditors only inspect about 8 GPs a month, whereas supervisors inspect 16 GPs a month.

Schargrodsky, 2003). Although this is a common anti-corruption policy prescription, there is little causal evidence that raising wages actually reduces corruption (Gans-Morse et al., 2018).

Honesty. We measure how higher salaries affect honesty by examining effects on (i) self-evaluations and (ii) inflation — our measure of the discrepancy between self-reported and supervisor-reported data for a subset of overlapping indicators. The inflation measure is described in detail in Section 3.3.1.

In Column (1) of Table 2, we observe that higher salaries caused self-reported scores to increase. Sub-Figure 2a shows how secretaries who receive the salary increase self-report very similar levels of performance before the salary increase, but those who received the salary increase start reporting a slightly higher performance in the months after the salary increase.[17] Appendix Sub-Figure A5a confirms this finding using a dynamic DiD approach. Column (1) in Table 2 shows that higher salaries cause self-reported scores to increase by 0.42 percentage points (or about 0.03 SD) in the 18 months after the salary increase, and this change is significant at the 1% level. Since we found no change in performance as measured by three different objective, hard-to-manipulate indicators — supervisor, auditor and citizen assessments — this suggests some increase in misreporting by GP secretaries. To probe this further, we examine how the salary increase affected inflation — our tight measure of misreporting for a subset of overlapping indicators tracked by both the secretaries and their supervisors.

Sub-Figure 2a presents time trends for inflation. We see no clear change after the salary increase. Column (2) of Table 2 shows that higher salaries have a negative, but statistically insignificant impact on inflation. Moreover, the effect size is also small: the point estimate implies that higher salaries caused inflation to reduce by 0.5pp, about 0.03 SD or 2.6% of the baseline mean. This fact suggests that self-reported scores do not change for overlapping indicators.[18] Appendix Table A5 confirms this: columns (2) and (4) show that higher salaries had no impact on self- or supervisor-reported scores for overlapping indicators.

[17] The self-reporting system was temporarily suspended between April 2022 - August 2022. We therefore have missing data across all GPs for these months.

[18] The performance indicators are described in detail in Appendix A.2.

The increase in self-reported scores and the absence of any change in inflation suggests that GP secretaries increased misreporting of scores for non-overlapping indicators that could not be verified by the supervisor. Columns (1) and (3) of Appendix Table A5 provide direct evidence for this: self-reported scores increase for non-overlapping indicators, even though supervisor assessments of those indicators show no change (if anything, a slight decrease). This effect is significant at the 1% level, but it is small in magnitude (about pp or 1.8% of the baseline mean).

Higher salaries increase the value of the job and may encourage secretaries to misreport data to inflate their performance. While the department was vague about how it would assess secretaries' performance when making tenure decisions and what would happen to secretaries who were not regularized, it was expected that supervisor evaluations and self-reported performance scores would both play a role. Thus, secretaries may have believed that inflating their self-reported scores boost their chances of regularization. Our results suggest that doubling bureaucrat salaries slightly increased misreporting, focused on hard-to-verify indicators, but this effect is small.

Corruption. As described in Section 3.3, we use two proxies of corruption. Our first proxy captures the extent of corruption or nepotism in local government hiring. Specifically, we measure the share of multi-purpose workers employed by the GP who have at least one name in common with the *sarpanch*, the GP's elected leader. Sub-figure 2d shows that treated and comparison GPs display parallel trends in this outcome before the policy change. We see no visible change in the share of same-name workers after the salary increase, suggesting no treatment effect on this type of corruption. This result is supported by the dynamic DiD estimates in Appendix Sub-Figure A5d. Column (3) of Table 2 shows the point estimate for the effect on this type of corruption is a statistically insignificant 0.168 percentage points change. Moreover, this null effect is precisely estimated: with 95% confidence, we can reject a change of more than 0.326 percentage points (0.01 SD).

Our second corruption proxy is the share of a GP's total expenditure that the secretary reimburses to himself. Sub-Figure 2c shows the time trends for this outcome. Again, we find no evidence that there was a divergence between the group of secretaries receiving the salary increase and the group that did not. Column (4) of Table 2 shows that the point estimate for the effect on this type of corruption is a statistically insignificant 0.00995 percentage points change. With 95% confidence, we can reject a change of more than 0.013 percentage points (0.011 SD).

Taken together, our evidence suggests that higher salaries had a limited impact on dishonesty or corruption in our context. This could reflect multiple factors: first, our measures of dishonesty and corruption may be imperfect and we may be missing some harder-to-detect forms of corruption. Second, the salary increase may be too small to affect corruption decisions on the margin. Third, the risk of detection and costs of being caught for corruption may still be too high or low so even a 91% increase in salaries does not change the cost-benefit calculation.

4.3 Selection and Retention

Retention. Even if the performance and honesty of incumbent bureaucrats are relatively inelastic to pay, raising salaries may be valuable if it helps to attract talented citizens to the bureaucracy and retain existing high-performing bureaucrats. We are unable to study how the salary increase impacted entry, since there has been no fresh recruitment of GP secretaries since the policy change. Thus, we focus on studying how salaries affect the retention of existing bureaucrats.

Figure 3a shows the effect of the salary increase on the share of secretaries who continue to be employed by the PRRD Department, even if they do not stay in exactly the same job. Appendix Figure A6 shows the dynamic DiD timelines corresponding to Sub-Figure 3a. We see parallel trends in stay rates before the salary increase, but over time salary increases reduced attrition among treated secretaries. This increase in stay rates does not appear in the first 5 months after the salary increase (by December 2021). However, 20 months after the salary increase (by March 2023), the stay rate is 2.2 percentage points higher among treated secretaries who received the increase. Column (1) of Table 3 confirms this result in a DiD regression. The effect size is substantial, amounting to a 1 percentage point reduction in quit rates per year (relative to a baseline of about 5% per year).

Composition of the Bureaucracy. Stay rates are important in their own right, as excessive turnover can be disruptive for organizations (Akhtari et al., 2022). However, the impact on public service delivery would be greater and more lasting if salaries affected the composition of the bureaucracy. We evaluate whether the salary increase had differential effects on the stay rates of secretaries based on their baseline performance and demographic characteristics.

First, we show in Sub-Figure 3b and Column (2) of Table 3 that the salary increase did not disproportionately affect retention of high-performing secretaries. We see no differential effect of the pay increase on the rates of secretaries with high baseline performance, as measured by supervisor evaluations.

Second, we test whether salaries are particularly effective at reducing attrition among particular groups of bureaucrats (e.g. women, disadvantaged castes). Focusing on gender, we observe that in May 2021 (the penultimate period before the policy change), only 29.2% of GP secretaries were female. Figure 3b and Appendix Table A6 shows that the overall positive effect of higher salaries on stay rates is primarily driven by male secretaries. The effect among women is 3.42 percentage points smaller for women and, when restricting the sample to only female secretaries, the baseline DiD coefficient is negative but not statistically significant. This might be because female secretaries may have fewer lucrative outside options that are female-friendly or considered socially appropriate for women. Turning to caste, we find that higher salaries do not differentially affect the stay rates of Scheduled Caste secretaries.

Thus, overall, our results suggest that higher salaries significantly reduce quit rates among bureaucrats, but do not drastically change the quality or composition of the bureaucracy.

4.4 Robustness Tests

Our main results are robust to a wide range of robustness tests. We have already shown that our results are robust to defining performance in a variety of ways. In addition, Appendix Table A7 shows that the results are robust to including linear and quadratic time trends interacted with the salary increase variable. Appendix Table A8 and Appendix Figures A7 and A8, show the effect of the salary increase on subjective assessments of the GP made by the supervisors, auditors, and citizens as opposed to the assessments constructed from the individual underlying indicators. Appendix Table A9 shows that the results are robust to including mandal-by-month fixed effects.[19]

Salary increases can affect public service delivery by (i) changing the behavior of incumbent bureaucrats or (ii) affecting the composition of the bureaucracy. In Appendix Table A10, we isolate the first channel by focusing on GPs where the same secretary was in charge throughout our study period. In this subsample, we again find that salary raises had no impact on bureaucrat performance, effort, dishonesty and corruption.

4.5 Expert Forecasts

Before releasing our results, we solicited predictions from forecasters on the Social Science Prediction Platform (SSPP) about the impacts of the salary increase. We described our context, policy change, and outcome variables, and asked the forecasters to predict the impacts of the salary increase on our key outcomes. Appendix E provides more information about the questions asked. We received a total of 85 predictions. Appendix Figure A9 provides an overview of the characteristics of the forecasters.

Performance. Table 4 shows the average and median predictions for the main outcomes among all respondents. Forecasters predicted that higher salaries would improve performance: the median and mean prediction is a 0.20 and 0.21 standard deviation increase in supervisor evaluation scores. By contrast, we are able to rule out improvements of more than 0.023 SD.

Corruption. The median prediction was that higher salaries would not change the level of nepotism in the hiring of multi-purpose workers, but would decrease reimbursement of GP funds by the secretary to themselves. By contrast, we find no impact on either outcome and are able to rule out effect sizes larger than 0.01 SD for the first outcome and 0.011 SD for the second outcome.

Retention and Selection. Forecasters accurately predicted that higher salaries would decrease attrition. The mean and median prediction is that the salary increase would reduce the number of secretaries leaving their jobs by 4.0 percentage points during the

[19]Mandals are smaller than districts, so this robustness test subsumes and is more demanding than district-by-month fixed effects, though our results remain similar when using those.

21 month period for which we have data, which is quite close to our finding (a 2.4pp decrease). While participants predicted that higher salaries would improve the average quality of secretaries by reducing attrition among top performers, we find no change in the composition of the bureaucracy.

Expertise, Contextual Knowledge and Accuracy of Predictions. 34.1% of the forecasters stated that they were "familiar" or "very familiar" with the academic literature on public sector pay while 12.9% were "familiar" or "very familiar" with the Indian context. 84.7% of the forecasters held a PhD degree or were in a PhD program. Appendix Table A12 shows that none of these groups have predictions that are substantially different from the overall sample.

5 Mechanisms, Relative Pay, and Heterogeneous Effects

Our results show that nearly doubling bureaucrat pay had (i) no average impact on performance and honesty and (ii) reduced attrition without significantly altering the composition of the bureaucracy. Yet these null *average* impacts could mask heterogeneous responses to the pay increase. In particular, higher salaries could trigger two offsetting forces: they could raise *morale*, but also weaken *incentives* by narrowing the gap between a bureaucrat's salary and her superior's pay. The strength of each channel could vary across PSes, and our reduced form results only capture their net effect.

5.1 Mechanisms: Morale and Incentives

To shed light on these underlying mechanisms, we examine treatment effect heterogeneity. Given our large sample of over 12,700 PSes, we are well-powered to test for heterogeneous effects. We identify groups of bureaucrats for whom the morale and incentives channels might be particularly strong or salient and estimate how they responded to the salary hike.

To investigate morale effects, we focus on bureaucrats who were dissatisfied with their employment conditions at baseline (14.6% of PSes) or explicitly complained about their salaries in an open-ended feedback form (17.4% of PSes).[20] We find that these bureaucrats

[20] We identify salary complaints using the following keywords: salaries, pay, wage, compensation, earnings, income, remuneration, package, benefits, bonus, and allowance.

respond more positively to pay increases by raising effort (figure A11). This pattern is particularly noticeable for PSes who had complained about their salaries.[21] These bureaucrats may have felt underpaid at baseline, and prior research suggests that workers may reduce effort if paid below their reference wage (Akerlof and Yellen, 1990; Mas, 2006). Pay increases when wages are below this reference point could boost effort, even if morale effects are weak for the average bureaucrat.

To explore the incentives channel, we identify PSes who were concerned about being regularized and promoted to match the pay and terms of Grade IV PSes. Such bureaucrats responded more negatively to pay increases by reducing effort. The coefficients in figure A11 show a consistent sign across outcomes though they are not always statistically significant. These results are consistent with the idea that higher salaries reduced Junior PSes' performance incentives, since the pay hike reduced the monetary returns to getting promoted.

5.2 Heterogeneous Effects: Traditional and GenericML Approaches

5.2.1 Traditional Approach

We estimate how treatment effects vary with bureaucrat demographic characteristics such as gender and caste, and other variables such as baseline performance.[22] Using the traditional approach, where we estimate a triple difference regression interacting $Post \times PayIncrease$ with secretary characteristics, we find limited evidence of treatment effect heterogeneity along these dimensions, as appendix figures A11 and A12 as well as appendix tables A13-A16 illustrate. However, the limited heterogeneity that we observe could reflect some of the limitations of the traditional approach, as different secretary characteristics may be correlated and interact with treatment in overlapping ways, masking the impact of any specific characteristic (Chernozhukov et al., 2018). Thus, we also employ a machine learning-based approach to estimating treatment effect heterogeneity, following the GenericML method of Chernozhukov et al. (2018).

[21] It is not unusual to observe changes in effort but no change in performance, since effort is often more elastic to pay and other morale changes than performance (DellaVigna et al., 2022).

[22] The baseline frequencies of these characteristics are as follows: 18.2% are of a scheduled caste, 29.2% are women.

5.2.2 GenericML Approach

Method. First, we train separate prediction models for the treatment and comparison groups. These models are then used to predict counterfactual outcomes for each observation, i.e. the outcome had a treated (comparison) observation with characteristics X_i been in the comparison (treatment) group. We can then obtain predicted treatment effects for each secretary, and present these for four key outcomes — performance (as measured by supervisor evaluations), effort, honesty, and stay rates. The GenericML procedure is described in more detail in Appendix B.

Predicted Treatment Effects. For each outcome, every secretary is assigned a quintile based on her predicted treatment effect. Appendix figures A13, A14, A15 display predicted treatment effects, grouped by quintile, which Chernozhukov et al. (2018) call Group Average Treatment Effects (GATES). There are two main takeaways from the GATES figures. First, for all four outcomes, most bureaucrats are predicted to have null treatment effects. Second, across all outcomes, we consistently see evidence of some heterogeneity, with the top quintile typically showing positive treatment effects while the bottom quintile shows negative treatment effects.

Treatment Effects Across Outcomes. We construct variables indicating whether a secretary is predicted to have a positive, null, or negative treatment effect for each outcome based on whether the 90% confidence interval of her predicted treatment effect lies fully above, fully below, or includes zero. In Appendix Table A17, we present evidence that treatment effects are correlated across outcomes: secretaries predicted to have positive treatment effects on supervisor evaluations are also predicted to raise effort and behave more honestly, i.e. reduce corruption (Column 1). Similarly, secretaries for whom higher salaries improve retention are also predicted to increase performance (Column 4). These correlations suggest that, for some secretaries, higher salaries had positive impacts on multiple outcomes, whereas, for other secretaries, the pay raise had no or even negative impacts across outcomes.

Appendix Table A18 shows the share of secretaries with different combinations of predicted treatment effects across outcomes. No secretaries are predicted to have positive or negative treatment effects on all four outcomes. In the second row, we identify secretaries for whom higher salaries are predicted to have weakly positive effects — improvements on some outcomes and no worse performance on any. About 21% of secretaries are in this category. For a smaller but still meaningful share (16%), salary increases are predicted to have weakly negative effects, worsening performance on some outcome and not improving any other. For a small subset (4%), higher salaries lead to tradeoffs—gains on some outcomes and worse performance on others. For the majority of secretaries (59%), we see null effects on all outcomes. This breakdown is consistent with our null reduced form results, but allows for a richer interpretation, given the offsetting heterogeneous responses of different bureaucrats.

Characteristics Associated with Treatment Effect Heterogeneity We describe the characteristics of secretaries in different GATES quintiles, which Chernozhukov et al. (2018) refer to as Classification Analysis (CLAN). Appendix figures A16, A17 and A18 show the characteristics that are most different between bureaucrats predicted to have the most positive vs the most negative treatment effects. Statistically significant differences are highlighted in blue. Secretaries who complained about their salaries are consistently predicted to respond more positively to pay hikes. These secretaries are predicted to have more positive treatment effects on supervisor evaluations, stay rates and effort, and reduced self-dealing (though this last difference is not statistically significant). Consistent with this, Appendix Table A18 shows that higher salaries have weakly positive impacts on a higher proportion (29%) of these secretaries and weakly negative impacts on a smaller proportion (11%) of such secretaries. This confirms our finding from the traditional heterogeneous effects analysis. Furthermore, the GenericML analysis does not identify substantial systematic heterogeneity for most other bureaucrat characteristics.[23]

[23] One exception are secretaries who perform well at baseline. The CLAN suggests that these secretaries respond more negatively to salary increases, as their effort and performance decrease. This could be because they interpreted the unconditional salary increase as a signal that performance would not be an important factor in regularization process or because the compression of the overall pay structure had an especially large disincentive effect on them.

5.3 Relative Pay Effects

Finally, having shown that raising Junior PSes' absolute salaries had no impact on their performance, we explore the effects of changes in relative pay, specifically pay rank, which prior work suggests may impact worker morale (Card et al., 2012; Breza et al., 2018). To identify the effects of pay rank, we focus on a group of bureaucrats whose absolute pay was unaffected by the reform, but whose rank in the pay hierarchy was. Grade IV PSes occupy the position in the salary scale just above Junior PSes, and were paid roughly twice as much as Junior PSes before the reform. The pay hike for Junior PSes equalized their salaries, placing both at the bottom of the new salary scale.[24] By contrast, the reform did not affect the pay rank of Regular PSes in the higher salary grades (I, II and III), who were more senior officers and were paid considerably more than Grade IV PSes.

We estimate a DiD design that compares the performance of (i) Grade IV vs other regular PSes (i.e. in salary grades I-III), (ii) before vs after the reform. As Appendix Figure A10 shows, we observe parallel trends between both groups for all outcomes, supporting the validity of this design. In addition, there is no visible change in trends of Grade IV PSes after the reform. This null result is corroborated by the DiD point estimates in Appendix Table A11. Taken together, these findings suggest that, at least in our context, changes in pay rank had no impact on bureaucrat performance or effort.

6 Conclusion

In this paper, we studied how bureaucrat salaries affect performance, effort, honesty and retention. We focused on GP secretaries, an important group of bureaucrats in India who are tasked with delivering a wide range of local public services and development programs — a broad role that resembles the job scope of core civil service employees. Our empirical strategy exploits a policy reform in Telangana, where salaries for one group of GP secretaries were doubled while salaries for other secretaries performing the same job

[24] Strictly speaking, a group of temporary PSes — called Outsourcing Panchayat Secretaries (OPS) — were the lowest pay, but they make up a small share of the sample. Their contractual arrangements are also different from the other PSes: the OPSes are technically employed by a contract agency and paid directly by that firm. Since OPSes pay and contract terms do not change before vs after July 2021, we include them in our baseline analysis and consider them along with "Regular PSes". Nevertheless, our results are all robust to dropping OPSes from the sample.

were left unchanged. This allows to use a DiD design to evaluate the impact of increasing bureaucrat salaries.

We find that salaries have a precisely estimated zero impact on bureaucrat performance. Higher salaries reduce effort slightly but do not impact the level of dishonest and corrupt behavior. The salary increase meaningfully improves retention, but otherwise does not change the quality or composition of the bureaucracy. Our results are at odds with the forecasts of experts, who predicted that higher salaries would improve performance, reduce corruption, lower quit rates and improve selection. While higher salaries often drive efficiency wage dynamics in the private sector (Coviello et al., 2022), we find no such evidence, which may reflect the comparatively low risk of termination in the public sector.

We found suggestive evidence that higher salaries affected both morale and incentives of lower-level bureaucrats. Higher salaries raise effort among bureaucrats who had complained about their pay, while depressing effort among bureaucrats who were concerned with getting promoted to the next pay grade. Prior work suggests that salary increases could engender reciprocity towards employers (Abeler et al., 2010; DellaVigna et al., 2022). However, if this was the primary mechanism, we would see effort increases across the distribution, and not only among bureaucrats with salary complaints. Instead, our results suggests that pay hikes increase effort when a bureaucrat is paid below her idea of a fair wage, consistent with the findings and framework of Mas (2006). This finding also highlights the dangers of reducing pay or setting salaries at very low levels. If overall salary levels are so low that many bureaucrats feel unfairly paid, increasing salaries may have had an overall positive effect on effort and performance.

Our results suggest that raising bureaucrat salaries is unlikely to impact performance or corruption — except under certain conditions. If most bureaucrats believe they are underpaid, or attrition is very costly, salary increases could pay dividends. However, in most developing countries, neither condition is likely to hold — lower-level bureaucrats are often paid well relative to their outside options, and attrition is typically low. A better understanding of what motivates bureaucrats and informs their idea of a fair wage can help us design remuneration systems that reward public service — and attract and retain individuals committed to this goal.

References

Abeler, J., S. Altmann, S. Kube, and M. Wibral (2010). Gift exchange and workers' fairness concerns: when equality is unfair. *Journal of the European Economic Association 8*(6), 1299–1324.

Akerlof, G. A. and J. L. Yellen (1990). The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics 105*(2), 255–283.

Akhtari, M., D. Moreira, and L. Trucco (2022). Political turnover, bureaucratic turnover, and the quality of public services. *American Economic Review 112*(2), 442–493.

Asher, S., T. Lunt, R. Matsuura, and P. Novosad (2021). Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the shrug open data platform. *The World Bank Economic Review 35*(4), 845–871.

Ashraf, N., O. Bandiera, E. Davenport, and S. S. Lee (2020). Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *American Economic Review 110*(5), 1355–1394.

Baicker, K. and M. Jacobson (2007). Finders keepers: Forfeiture laws, policing incentives, and local budgets. *Journal of Public Economics 91*(11-12), 2113–2136.

Barteska, P. (2024). Bureaucrats and the korean export miracle. *Working Paper*.

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

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.

Besley, T., R. Burgess, A. Khan, and G. Xu (2022). Bureaucracy and development. *Annual Review of Economics 14*, 397–424.

Best, M. C., J. Hjort, and D. Szakonyi (2023). Individuals and organizations as sources of state effectiveness. *American Economic Review 113*(8), 2121–2167.

Breza, E., S. Kaur, and Y. Shamdasani (2018). The morale effects of pay inequality. *The Quarterly Journal of Economics 133*(2), 611–663.

Cabrera, J. M. and D. Webbink (2020). Do higher salaries yield better teachers and better student outcomes? *Journal of Human Resources 55*(4), 1222–1257.

Card, D., A. Mas, E. Moretti, and E. Saez (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review 102*(6), 2981–3003.

Census of India (2011). Census of India.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. Technical report, National Bureau of Economic Research.

Chi, G., H. Fang, S. Chatterjee, and J. E. Blumenstock (2022). Microestimates of wealth for all low-and middle-income countries. *Proceedings of the National Academy of Sciences 119*(3), e2113658119.

Coviello, D., E. Deserranno, and N. Persico (2022). Minimum wage and individual worker productivity: Evidence from a large US retailer. *Journal of Political Economy 130*(9), 2315–2360.

Dal Bó, E., F. Finan, and M. A. Rossi (2013). Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service. *The Quarterly Journal of Economics 128*(3), 1169–1218.

de Ree, J., K. Muralidharan, M. Pradhan, and H. Rogers (2017). Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia*. *The Quarterly Journal of Economics 133*(2), 993–1039.

De Ree, J., K. Muralidharan, M. Pradhan, and H. Rogers (2018). Double for nothing? Experimental evidence on an unconditional teacher salary increase in Indonesia. *The Quarterly Journal of Economics 133*(2), 993–1039.

DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2022). Estimating social preferences and gift exchange at work. *American Economic Review 112*(3), 1038–1074.

Demirgüç-Kunt, A., M. Lokshin, and V. Kolchin (2023). Effects of public sector wages on corruption: Wage inequality matters. *Journal of Comparative Economics 51*(3), 941–959.

Deserranno, E., P. Kastrau, and G. León-Ciliotta (2024). Promotions and productivity: the role of meritocracy and pay progression in the public sector. *American Economic Review: Insights* (Forthcoming).

Di Tella, R. and E. Schargrodsky (2003). The role of wages and auditing during a crackdown on corruption in the city of Buenos Aires. *The Journal of Law and Economics 46*(1), 269–292.

Duflo, E., R. Hanna, and S. P. Rya (2012). Incentives work: Getting teachers to come to school. *The American Economic Review 102*(4), 1241–1278.

Finan, F., B. A. Olken, and R. Pande (2017). The personnel economics of the developing state. *Handbook of economic field experiments 2*, 467–514.

Foltz, J. D. and K. A. Opoku-Agyemang (2015). Do higher salaries lower petty corruption? a policy experiment on west africa's highways. *Working Paper 12*, 245.

Gans-Morse, J., M. Borges, A. Makarin, T. Mannah-Blankson, A. Nickow, and D. Zhang (2018). Reducing bureaucratic corruption: Interdisciplinary perspectives on what works. *World development 105*, 171–188.

George, S., V. Rao, and M. R. Sharan (2024). Two Hundred and Fifty-Thousand Democracies. *Development Research*.

Hanna, R. and S.-Y. Wang (2013). Dishonesty and Selection into Public Service. Technical report, National Bureau of Economic Research.

Hasnain, Z., N. Manning, and J. H. Pierskalla (2014). The promise of performance pay? Reasons for caution in policy prescriptions in the core civil service. *The World Bank*

Research Observer 29(2), 235–264.

Henderson, J. V., A. Storeygard, and D. N. Weil (2011). A Bright Idea for Mesuring Economic Growth. *American Economic Review*.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *American Economic Review 102*(2), 994–1028.

Holmstrom, B. and P. Milgrom (1991). Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *The Journal of Law, Economics, and Organization 7*(special_issue), 24–52.

Iyer, L. and A. Mani (2012). Traveling agents: political change and bureaucratic turnover in India. *Review of Economics and Statistics 94*(3), 723–739.

Jayaraman, R., D. Ray, and F. De Véricourt (2016). Anatomy of a contract change. *American Economic Review 106*(2), 316–358.

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), 237–270.

Khan, M. Y. (2023). Mission Motivation and Public Sector Performance: Experimental Evidence from Pakistan. Working paper.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Lehne, J., J. N. Shapiro, and O. V. Eynde (2018). Building connections: Political corruption and road construction in India. *Journal of Development Economics 131*, 62–78.

Mas, A. (2006). Pay, reference points, and police performance. *The Quarterly Journal of Economics 121*(3), 783–821.

Mattsson, M. (2024). Information systems, service delivery, and corruption: Evidence from the bangladesh civil service. *Working Paper*.

Muralidharan, K. and V. Sundararaman (2011). Teacher Performance Pay: Experimental Evidence from India. *The Journal of Political Economy 119*(1), 39–77.

Niehaus, P. and S. Sukhtankar (2013). Corruption Dynamics: The Golden Goose Effect. *American Economic Journal: Economic Policy 5*(4), 230–269.

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

Rao, V., K. Ananthpur, and K. Malik (2017). The anatomy of failure: An ethnography of a randomized trial to deepen democracy in rural india. *World development 99*, 481–497.

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.

Rauch, J. E. and P. B. Evans (2000). Bureaucratic structure and bureaucratic performance in less developed countries. *Journal of public economics 75*(1), 49–71.

Socio Economic and Caste Census (2011). Socio Economic and Caste Census.

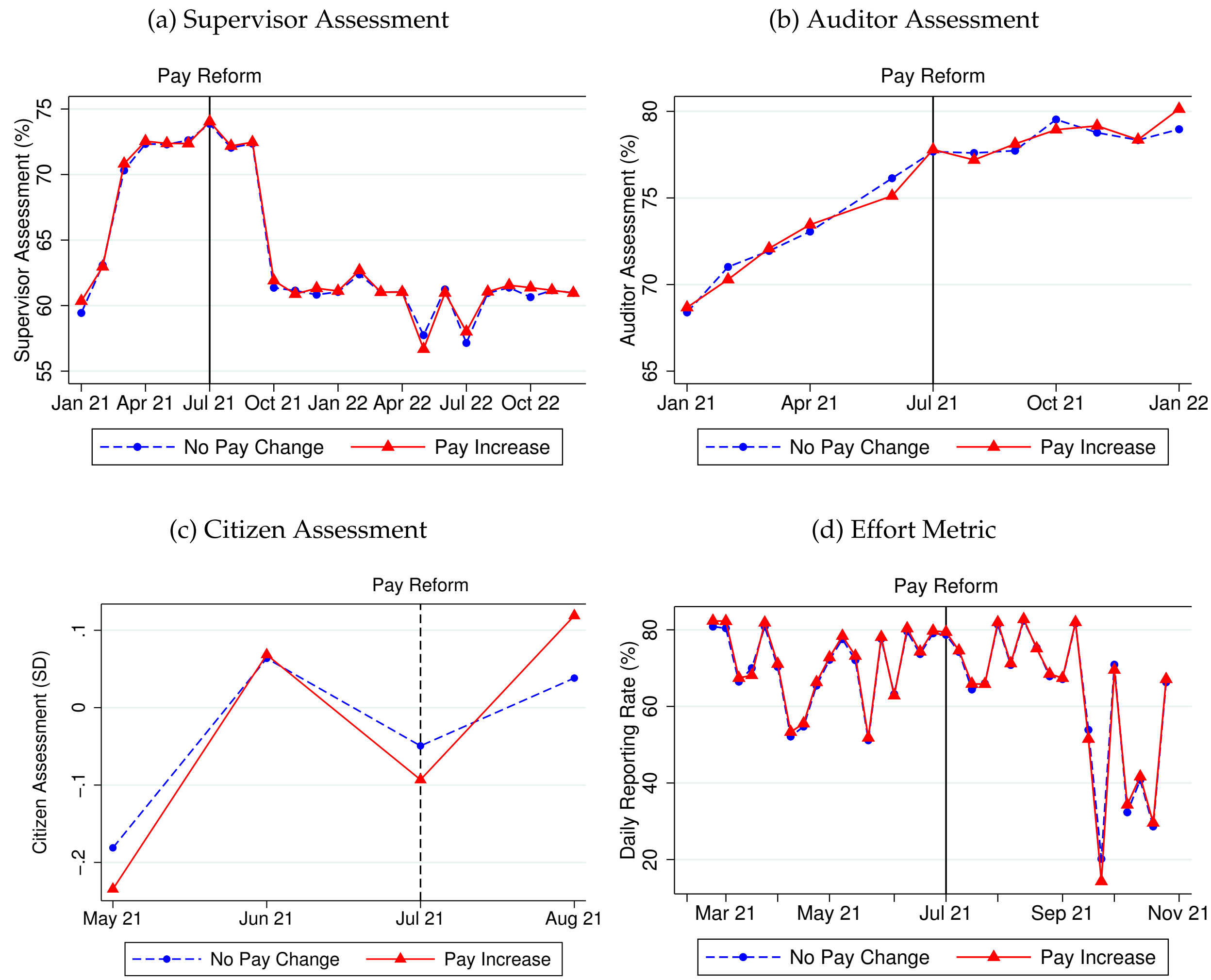
Van Rijckeghem, C. and B. Weder (2001). Bureaucratic corruption and the rate of temptation: do wages in the civil service affect corruption, and by how much? *Journal of development economics 65*(2), 307–331.

Verma, R. and N. Ranjan (2023). Panchayat secretary: The last-mile bureaucrat. *Working Paper*.

Willén, A. (2021). Decentralization of wage determination: Evidence from a national teacher reform. *Journal of Public Economics 198*, 104388.

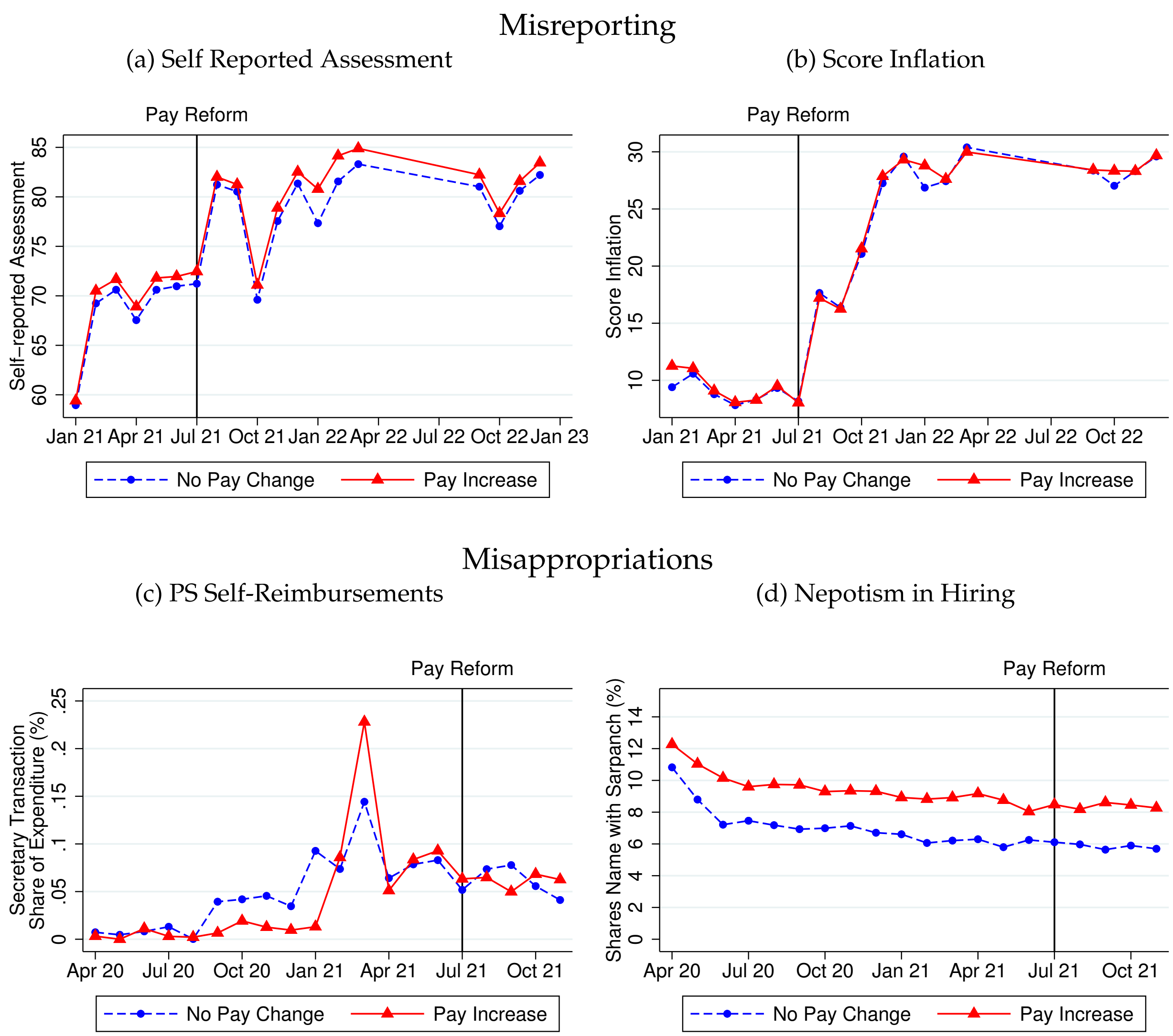
World Bank (2023). World Bureaucracy Indicators. Database.

Figure 1: Effect of Salary Increase on Performance



Note: This figure presents visualizations of the effect of the salary increase on secretary performance. We present averages of minimally processed data for the four performance outcome variables. In each figure, we present monthly averages for treated GPs (whose secretaries receive the salary increase) and comparison GPs (whose secretaries do not). Sub-Figure (a) shows timelines for the performance scores calculated from the supervisor assessments. Sub-Figure (b) shows the timelines for performance scores calculated from state-level auditors. Sub-Figure (c) shows the timelines for the performance scores calculated from the citizen assessments. Sub-Figure (d) shows how regularly the secretary files the required daily sanitation report, which we use as a proxy for effort. In both sub-figure (a) and (b) the outcomes are expressed as percentage out of 100. The outcome in Sub-figure (c) is expressed as SDs from the mean. The outcome in Sub-figure (d) is expressed as the percentage of total working days. Table 1 shows the results from DiD regressions using these outcome variables. See discussion in Section 4.1.

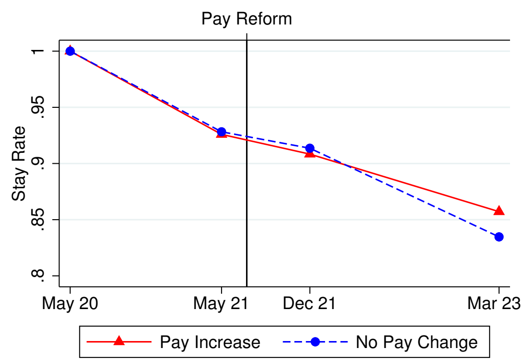
Figure 2: Effect of Salary Increase on Honesty and Corruption



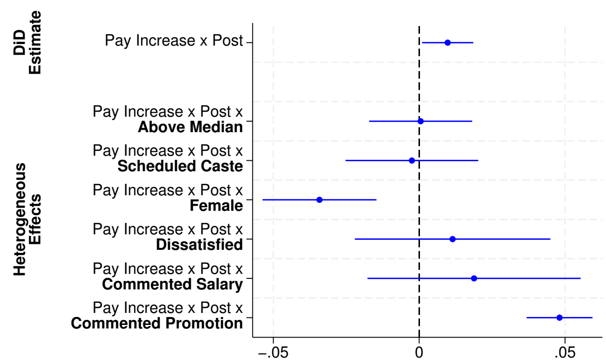
Note: This figure presents visualizations for the effect of the salary increase on honesty and corruption measures. We present time trends based on averages of minimally processed data for secretaries self-reported performance, score inflation, and two proxies for corruption. Sub-Figure (a) shows the timelines for the secretaries' self-reported assessment expressed as percentage out of 100. Sub-Figure (b) shows secretaries' score inflation, as defined in Section 3.3.1, expressed as a percentage. Sub-figure (c) shows the share of GP expenditure that was spent in reimbursing the GP secretary for personal expenses, expressed as a percentage of total monthly expenditure. Sub-Figure (d) shows the changes over time for the share of multipurpose workers employed by the GP who share a name with the GP Sarpanch, expressed as a percentage of all multipurpose workers hired by that GP in a month. The vertical lines indicate July 2021, the month of the salary increase. Table 2 shows the results from DiD regressions using these outcome variables. Appendix Figure A5 shows the dynamic DiD timelines for these outcomes. See discussion in Section 4.2.

Figure 3: Effect of Salary Increase on Retention

(a) Average Impact on Secretary Stay Rates



(b) Heterogeneous Effect



Notes: This figure shows how higher salaries affect retention, i.e. whether GP secretaries stay in government service. Sub-Figure (a) shows the share of GP secretaries holding a position in April 2020 who are still employed by the PRRD Department (even in another position) in March 2023. Sub-Figure (b) shows a coefficient plot of DiD estimates (with 95% confidence intervals). Each coefficient comes from a different regression. The first coefficient from the top is the main DiD estimate capturing the average effect of higher salaries on stay rates. The subsequent coefficients represent the triple interaction between *Pay Increase*, *Post* and the respective GP secretary characteristic. *Above Median* is a dummy for whether the secretary has a baseline supervisor evaluation above the median, *Dissatisfied* indicates whether the secretary reported by unhappy or very unhappy about work in a survey in May 2021. *Commented Salary* indicates whether the secretary made specific complaints about salary-related problems in the May 2021 survey. *Promotion* indicates whether the secretary made specific complaints about promotion-related problems in the May 2021 survey (the interaction term *Post* × *Commented Promotion* is excluded due to collinearity).Table 3 shows the DiD estimate of Sub-figure (a) as well as the difference in staying rate for those who were above the median at baseline. Table A6 shows the heterogeneous DiD regression estimates. See discussion in Section 4.3.

Table 1: Impacts of Salary Increase on Secretary Performance — Supervisor Score, Auditor Score, Citizen Score, and Effort

	(1) Supervisor Assessment	(2) Auditor Assessment	(3) Citizen Assessment	(4) Effort Metric
Pay Increase x Post	-0.115	0.171	0.0224	-0.707***
	(0.231)	(0.566)	(0.0726)	(0.174)
Observations	105,732	12,221	9,512	459,720
Clusters	12,678	4,851	3,114	12,770
Time FE	Yes	Yes	Yes	Yes
Panchayat FE	Yes	Yes	Yes	Yes
Dep. Var. Mean	64.7	75.7	.019	66.634

Notes: This table shows the estimated effects of the salary increase on the performance of the secretaries, displaying the DiD coefficient β from estimating Equation (1). In Column (1) the outcome variable is the performance score constructed from the supervisor evaluations, in Column (2), it is the performance score constructed from auditor evaluations. For Column (3), it is the score constructed from citizen assessments, and in Column (4) it is the daily reporting rate, i.e. how regularly secretaries file the required daily sanitation report, which is a proxy for effort. Standard errors are clustered at the GP level. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.1.

Table 2: Effect of Salary Increase on Honesty and Corruption

	(1) Self-reported Assessment	(2) Score Inflation (%)	(3) Sarpanch Self-Dealing (%)	(4) Secretary Self-Reimbursements (%)
Pay Increase x Post	0.417***	-0.499	0.168	0.00995
	(0.141)	(0.317)	(0.252)	(0.0118)
Observations	242,630	89,239	139,045	247,766
Clusters	12,770	12,639	10,503	12,770
Time FE	Yes	Yes	Yes	Yes
Panchayat FE	Yes	Yes	Yes	Yes
Dep. Var. Mean	76.3	18.99	8.44	.047

Notes: This table shows the estimated effects of the salary increase on self-reported performance and score inflation by the GP secretaries as well as two proxies for corruption. We estimate Equation (1) and display the DiD coefficient β. In Column (1) the outcome variable is the performance index constructed from the self-reported performance indicators, expressed as a percentage out of 100. In Column (2) the outcome variable is score inflation, the discrepancy between self-reported scores and supervisor evaluations for overlapping indicators. As defined in Section 3.3.1, the differences are aggregated into an index and expressed as a percentage out of 100. In Column (3), the outcome variable is the share of multi-purpose workers employed by the GP who share a name with the GP Sarpanch. In Column (4), the outcome is the share of GP expenditure that was spent in reimbursing the GP secretary for his or her personal expenses, expressed as the percentage of total monthly expenditure. Standard errors are clustered at the GP level. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.2.

Table 3: Effect of Salary Increase on Retention

	(1) Employed	(2) Employed	(3) Employed
Pay Increase x Post	0.00976**		0.00307
	(0.00448)		(0.00635)
Pay Increase x Dec 21		-0.00408	
		(0.00376)	
Pay Increase x Mar 23		0.0236***	
		(0.00642)	
Pay Increase x Post x Above Median Score			0.000501
			(0.00901)
Post x Above Median Score			0.00503
			(0.00779)
Observations	45,724	45,724	44,180
Clusters	11,431	11,431	11,045
Time FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Dep. Var. Mean	.922	.922	.924

Notes: This table shows the estimated effects of the salary increase on retention, i.e. whether GP secretaries remain public sector employees. We also examine whether effects vary with baseline performance, as measured by supervisor evaluations. In all columns, the outcome variable is a dummy indicating whether a secretary is still employed in the PRRD Department, even if in another position. Standard errors are clustered at the employee level. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.3.

Table 4: Expert Forecasts

Overall Predictions			
Outcome	Median	Mean	Std. Dev.
Supervisor Assessment	0.20	0.21	(0.19)
Self-reported Assessment	0.30	0.33	(0.21)
Stay-rate	4.08	3.98	(2.93)
	Share of Predictions (%)		
	Worsen	Unchanged	Improves
Selection	27.10	14.10	58.80
	Decreases	Unchanged	Increases
Secretary Corruption Proxy	74.10	15.30	10.60
Sarpanch Corruption Proxy	47.10	32.90	20.00
Observations	85		

Notes: This table presents summary statistics of the forecasters predictions. Forecasters were recruited via the Social Science Prediction Platform. Predictions for supervisor assessment and self-reported assessment are in SDs from the mean. Predictions for stay rates are in percentage points. See discussion in Section 4.5

Appendix for Online Publication

A Data Appendix

A.1 Roster of Public Sector Employees

Data on which secretary held which office at what time, is constructed from four listings of all the employees in the GP and Mandal public servants. These were published in May 2020, May 2021, December 2021, and March 2023. Each listing has an employee ID, which allows us to track employees when they are transferred from one GP to another or if they were transferred from a secretary position to a different role within the PRRD Department. We use this to create a panel dataset with 4 time periods for each secretary.

We determine the gender of each secretary using the 2020 data on the employee's gender. Caste information comes from the Mar 2023 data as well as an incomplete dataset from April 2020. These variables are used for the heterogeneity analysis described in Section 5.

A.2 Performance Indicators

A.2.1 Supervisor Score

We construct supervisor evaluation scores using data from supervisor inspections. Following the government aggregation of the data, we create 10 components, each representing a key responsibility of the secretary. As supervisors sometimes do not fill out all sub-scores, we create a percentage for each category by dividing the total score by the maximum possible score among the non-missing sub-scores. The scores are aggregated to the GP-month level. For details about each of the components and how they are aggregated, see Table A3.

A.2.2 Self-reported Performance

Self-reported performance scores are constructed from data that is self-reported by GP secretaries. The self-reported indicators are aggregated by the government score into 12 categories. We then apply the government's weights across categories to construct a weighted average total score, and normalize it as a percentage out of 100. The scores are available at the GP-month level. The categories in the self-reported score are the same as those in the supervisor evaluations, except that it additionally includes a National Rural Employment Guarantee Act (NREGA) score, which tracks NREGA social welfare scheme implementation by the secretary. The self-reported indicators generally have high overlap with the supervisor-collected indicators, but there are some which cannot be verified by the supervisor (e.g. self-declared frequency of road and drain cleaning). For details about each of the components and how they are aggregated, see Table A3.

A.2.3 Score Inflation

Each supervisor assessment component has a corresponding self-reported assessment component. We create a separate secretary score that is used for inflation calculations, creating sub-components that are analogous to the supervisor score components using inspection data. This allows us to examine the inflation of specific secretary score components. For a given month and year, we calculate inflation only for the components that overlap between the secretary and supervisor scores. We determine the inflation percent by calculating the absolute difference in the overlapping components, and then express this as a percentage point of the total possible score. This percentage represents the amount by which the secretary's score exceeds or falls short of the supervisor's score. The inflation formula is described in Section 3.3.1.

A.2.4 Bureaucrat Effort

As a proxy for effort, we measure the reporting rate of secretaries. Secretaries are required to submit a daily sanitation report. We create a daily panel with a dummy variable indicating whether each secretary submitted a report for each working day of 2021. This covers a period 20 weeks before and after the policy of salary increase took place. This panel was then averaged at the GP-week level. Due to the daily reporting requirement, the effort proxy is at a higher frequency than the other performance measures.

A.2.5 State-level Auditor Score

State-level scores are calculated from sub-scores that state-level auditors give on their visits to GP (which happens on average of thrice a year for any given GP). These ratings are scores given out of 10, for the quality or cleanliness of roads, drains, institutions, street lights, plantations, plant nurseries, waste segregation and garbage segregation. These scores are individually converted to percentages out of 100 and then averaged to create the state-level auditor score. Finally, they are averaged to create a panel at the GP-month level. For details about each of the components and how they are aggregated, see Table A3.

A.2.6 Citizen Score

The third performance indicator is a citizen survey. 10,606 citizens were surveyed in a total of 4,672 GPs (about 25% of all GPs) over 4 months, covering a period before and after the pay change (May-Aug 2021). These surveys took place via phone and were conducted by a survey company contracted by the PRRD Department. GPs were first randomly sampled from all mandals. Then, within each sampled GP, citizens were randomly sampled based on phone numbers in the Dharani and Public Distribution System database, which maintain property records and beneficiary lists for subsidized food respectively. These databases together cover over 90% of rural households.

Citizens were asked to rate public services using a Likert scale (Very poor, poor, average, good, Excellent). These sub-scores were converted to numeric values on a 1 to 5 scale, standardized, averaged, and then standardized again (i.e. the index construction procedure of Kling et al. (2007)) to form a citizen z-score at the citizen level. For details about each of the components and how they are aggregated, see Table A3.

A.2.7 Subjective Scores for Secretaries by Supervisors

Supervisors were also asked for their overall opinion on secretaries on a 0 to 5 scale. These scores were averaged on a village-month level, and then standardized. As these score are less objective in how they measure the performance and less comparable across supervisors we use them as a secondary outcome. We show the results using these subjective scores as the outcome variables in Appendix Figure A4 and Appendix Table A4.

A.2.8 Overall Scores for GPs by Supervisors, State-Level Auditors, and Citizens

In the same visits by supervisors and state-level auditors, they were asked about their general opinion of the village. Supervisors gave scores on a 0 to 5 scale, which was averaged on a GP-month level, and then standardized. State-level auditor scores were scores out of 10, which was converted to percentage out of 100%, and then averaged on a GP-month level. Citzens also gave their subjective opinion on their overall satisfaction of their village in the survey above, these scores were averaged at the GP-month level and then standardized. Appendix Table A8, Appendix Figures A7, and A8, show results for these outcome variables.

A.2.9 Baseline Scores for heterogeneity

The averages of the supervisor scores from the months before July 2021, are used to classify an above or below-median performing village secretary, which is then used in the heterogeneity analysis in 5. We supervisor evaluations to establish baseline performance, as these have the widest coverage of GPs.

A.3 GP Secretary Satisfaction

The survey about the Panchayat secretaries work conditions and job satisfaction took place between April-May 2021. 9,474 out of the 12,770 secretaries responded. Of the respondents, 7,114 (75.1%) were treated secretaries who later received the pay increase and 2,360 (24.9%) were comparison secretaries whose salary remained unchanged. Of the 9,474 respondents, 7,098 (74.9%) left a free text comment as a response. Of the comments mentioning salary, 87.5% were by treated secretaries (who initially received the lower salary and would later receive the pay increase) and 12.5% were comparison secretaries (who were already receiving the higher salary). It is not surprising that treated secretaries were more likely to make salary complaints since they were paid less at the time of the survey.

This survey also asked secretaries about their general morale and job satisfaction. We construct an indicator for whether a secretary was dissatisfied with their job and whether they made a comment about salary. This information is used in the heterogeneity analyses in Section 5. We classify a secretary as dissatisfied if they choose the bottom two options of the satisfaction scale (which contains options very dissatisfied, dissatisfied, neutral, satisfied, very satisfied). Secretaries would have been classified as commented about salary in the free text response in the survey if their complaints mentioned: "salaries", "pay", "wage", "compensation", "earnings", "income", "remuneration", "package", "benefits", "bonus", "allowance", "pay scale". Secretaries would have been classified as commented about promotion in the free text response in the survey if their complaints mentioned: "regularisation", "regularised", "regularized", "regularization", "regular ps", "job security".

A.4 Pay Increase in Net Present Value Terms

Many Junior PSes may have expected to be regularized as Grade IV secretaries at the end of their contract. Thus, while the policy change led to an immediate 91% increase in Junior PS' salaries , the expected net present value (NPV) of their labor income may have increased by a smaller amount. To estimate how the policy change affected the NPV of their income, we make several assumptions. First, we assume secretaries expect to hold the job for 20 years and discount future income over this period using an annual discount rate of 10%.[25] Furthermore, we assume that Junior PSes hold beliefs about their probability of being regularized in July 2023, approximately 4 years after they were hired. If a PS is not regularized, we assume that they would continue to work as a Junior PS for the rest of their 20-year tenure.

Under these assumptions, if Junior PSes believed there was no chance they would be regularized, the 2021 policy change would raise the NPV of their labour income stream by 91%. On the other hand, if they were certain that they would be regularized, the 2021 salary increase would have increased the NPV of their labour income by only 11%. If their beliefs were in-between these extremes (i.e. a 50% probability of being regularized), the salary increase would increase the NPV of their income by 40%.

While we have no explicit data on PSes' expectations about regularization, it is clear that regularization was far from guaranteed. Before entering service, all Junior PSes signed an Agreement Bond affirming that they "will not have any claim or right to be

[25]This discount rate is 3 percentage points higher than India's 20-year bond yield in July of 2021 but 1 to 4 percentage points lower than returns on the S&P Bombay Stock Exchange Index in the 5, 10, or 20 years preceding the salary change.

appointed on a regular basis and such appointment will be made without prejudice to the regular appointments that may be made in future." Moreover, this contract explicitly stated that "regularization of service [...] cannot be for all, but will be subject to the assessment and evaluation of their performance by a committee appointed by Government and those whose performance is found to be satisfactory can only be regularized." As late as April 2023, the PRRD Department sent a circular to all Junior PSes reminding them of their contractual terms. This was not just cheap talk: in a separate instance, the PRRD Department did not renew the contracts of many employees tasked with implementing the National Rural Employment Guarantee Scheme (NREGS), an important social protection programme, due to substandard performance.

Eventually, after their 4-year contract was completed, the Junior PSes went on strike to force the government's hand. In response, the PRRD Department created a committee to determine regularization criteria. Junior PSes who were assessed to perform well during their contract period — according to a variety of metrics — were regularized as Grade IV PSes. Overall, our understanding is that approximately 90% of Junior PSes were regularized in late 2023. Furthermore, as shown in Figure 3a, approximately 5% of PSes left the service each year. The fact only 90% of PSes were regularized and about 10% of them left service even before they could be regularized suggests that 80% is a reasonable expectation about whether a Junior PS would subsequently draw a Grade IV PS salary. With a prior belief of 80%, the pay hike would raise the NPV of their income by 29%. The NPV increase would be even higher if Junior PSes were hyperbolic discounters, who placed greater utility weight on present income or lacked trust that the government would deliver on their agreement.

Overall, these calculations suggest that the Junior PS pay hike that we study led to a significant increase in the NPV of the future income stream from holding the job as Junior PS.

A.5 Misappropriation data

Every GP keeps a record of all financial expenditures. From these transaction data, we generate the following measures of misappropriation at the GP-month level.

A.5.1 Share of Multi-Purpose Workers with the Same Name as Sarpanch

This is the share of the number of multi-purpose worker expense transactions where the worker's name contains a part of the Sarpanch's name. Transactions are tagged as being to multi-purpose workers if the description included "mpw", "multiple", "multi" or "mult". This variable varies at the GP-month level. Following Lehne et al. (2018), we consider this a proxy for nepotism in GP hiring. This is certainly an imperfect measure, since certain names may be more common and some GPs may be more ethnically or culturally homogeneous. However, these biases are most likely to be time-invariant. Since our empirical strategy is a DiD design, these time-invariant biases should be netted out. While our results remain vulnerable to time-varying measurement biases, we believe these biases are less obvious (at least it is harder to construct obvious reasons for these) and we also see no clear time trends in our data.

A.5.2 Secretary Reimbursement Share

This is the monthly share of the value of all expenditures that are for the purpose of reimbursing the secretary for personal expenses in a given month. Transactions are tagged as secretary transactions if the expense description contains the mention of "secretary". This is also a GP-month level variable.

A.6 SHRUG Data

The Socioeconomic High-resolution Rural-Urban Geographic Platform for India (SHRUG) (Asher et al., 2021) was used to obtain baseline covariates for both descriptive statistics in A2 as well as for testing for heterogeneous effects 5. GP latitude and longitude coordinates were merged with the SHRUG Shapefiles using a spatial join to match each GP to their respective SHRID polygon. Following this, using the SHRID identifiers, multiple datasets were merged in including 2011 population census data (Census of India, 2011), 2011-2012 socio-economic and caste census (Socio Economic and Caste Census, 2011), relative wealth estimates predicted by facebook data (Chi et al., 2022), and levels of night lights (Henderson et al., 2011). Furthermore, public goods indices for education, health, sanitation, connectivity, and overall public goods are calculated by scaling relevant variables and taking their means. The Education Index included variables related to various educational institutions, while the Health Index considered multiple healthcare facilities. The Sanitation Index accounted for sanitation and water facilities, and the Connectivity Index covered infrastructure like roads, communication, and transport services.

B GenericML Procedure

Since our baseline DiD specification includes GP and month fixed effects, we demean our outcome variables (supervisor score, staying rate, etc) to mimic the DiD specification. For the machine learning implementation, we specified a random forest model with 128 trees and applied the GenericML function with 100 sample splits. The propensity score was based on the share of secretaries who received the pay increase (approximately 71%). The predictors used in our model included the baseline supervisor scores, self-assessed scores, reporting rates, gender and caste of secretaries, dummies for baseline dissatisfaction and salary complaints, and the SHRUG covariates displayed in Table A2 (e.g. population, public goods indices, consumption, wealth, etc).

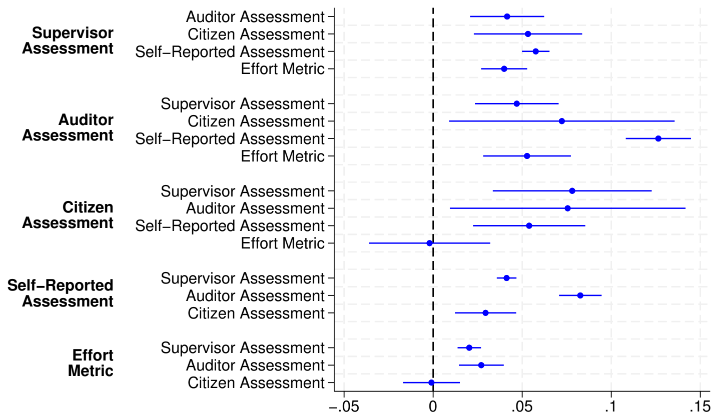
For each sample split, we first randomly split the data into two disjoint sets, Sample A and Sample M. In Sample A, we train machine learning models (in our case, random forest models) to obtain estimates B(Z) and S(Z), which correspond to the baseline conditional average $b_0(Z)$ (expected value of Y given Z when the treatment D is absent) and the conditional average treatment effect $s_0(Z)$ (captures the difference in the expected value of Y between those who receive the treatment (D=1) and those who do not (D=0), given Z), respectively.

We focus on estimating and making inferences about the key features of $s_0(Z)$ in Sample M. Specifically, for each secretary in Sample M, we calculate the predicted treatment effect by comparing the observed outcome in her actual group (e.g., treatment) with the predicted counterfactual outcome had she been in the other group (e.g., control). This difference provides a predicted treatment effect for each individual. Since there are two main sources of uncertainty—estimation uncertainty from the model training on Sample A and splitting uncertainty from the initial data partitioning—we repeat this process multiple times across 100 different splits to obtain more reliable estimates of the key features of the treatment effect. Taking the median predicted treatment effect for each secretary across splits, we then group secretaries into quintiles, defining cutoffs at the 20th, 40th, 60th and 80th percentiles. We then plot the distribution of the Sorted Group Average Treatment Effects (GATES). The confidence interval for each quintile are the 5th and 95th percentile of treatment effects, when the sample is restricted to this quintile and the baseline DiD regression is estimated on it.

Having assigned each secretary a quintile based on their predicted treatment effect, we conduct a Classification Analysis (CLAN), comparing the characteristics (i.e. predictor variables) of secretaries in the 5th quintile vs the 1st quintile. This tells us the difference between secretaries predicted to have the most positive treatment effects and the most negative treatment effects. The CLAN figures present these differences in coefficient plot form, with blue shading indicating that the group difference for that characteristic is statistically significant in at least 90% of sample splits.

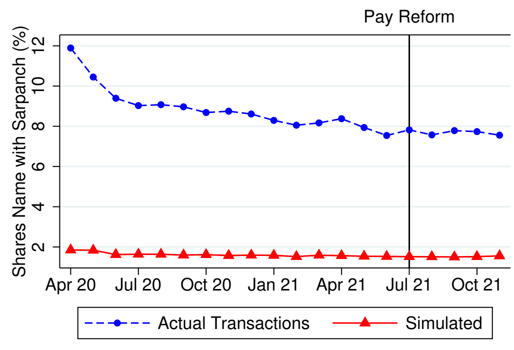
C Additional Figures

Figure A1: Coefficient Plot of Pairwise Regression of Main Outcome Variables



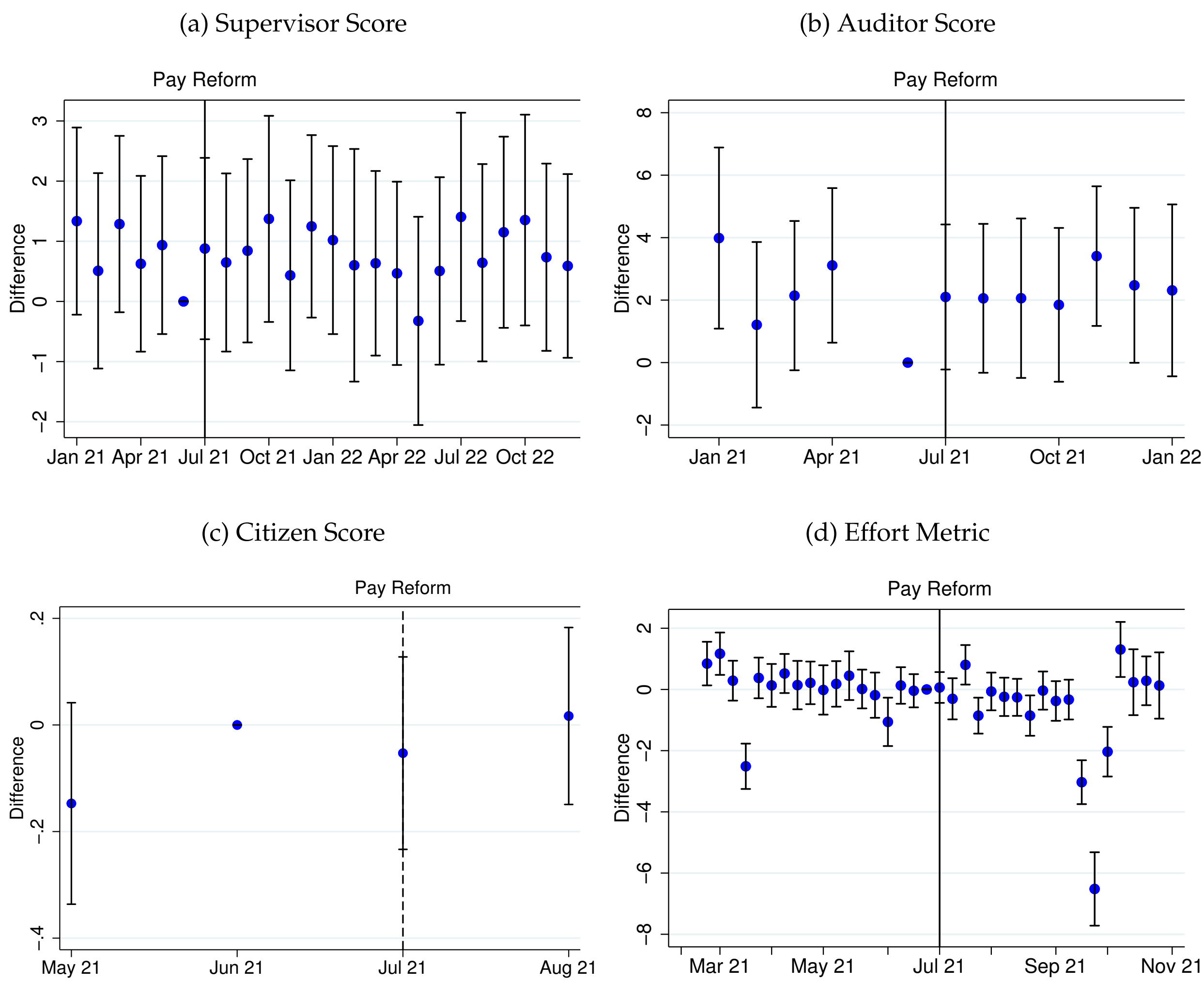
Notes: This figure illustrates the correlation between our main performance variables. This figure presents beta coefficients from pairwise regressions involving five variables: supervisor assessments, auditor assessments, citizen assessments, self-reported assessments, and our proxy for effort. For each variable, four regressions were conducted, treating it as the dependent variable and each of the other three as the independent variable. For example, the first regression from the top is estimated as $SupervisorAssessment_{it} = \alpha + \beta \times AuditorAssessment_{it} + \tau_t + \varepsilon_{it}$, with analogous models for the other permutations. The figure illustrates the beta coefficients from these regressions along with their 95% confidence intervals constructed using standard error clustered at the GP level. To ensure comparability among the groups, all variables have been standardized to have a mean of zero and standard deviation of one before the regressions. As the self-reported assessment and the effort metric are mechanically correlated we do present these coefficients, although they are large and statistically significant. See discussion in Section 3.

Figure A2: Sarpanch and Worker with the Same Name: Actual vs Benchmark



Notes: This figure shows the share of multi-purpose workers employed by the GP who share a name with the Sarpanch. The blue dashed line shows actual transactions. The red solid line shows transactions where, instead of using the actual Sarpanch's name, we selected a random sarpanch name from our sample. The solid red line illustrates the rate at which the Sarpanch and multipurpose workers would have the same name if the GP hired multipurpose workers at random from the citizenry and there was no nepotism or corruption. See discussion in Section 3.3

Figure A3: Effect of Salary Increase on Performance



(a) Supervisor Score

(b) Auditor Score

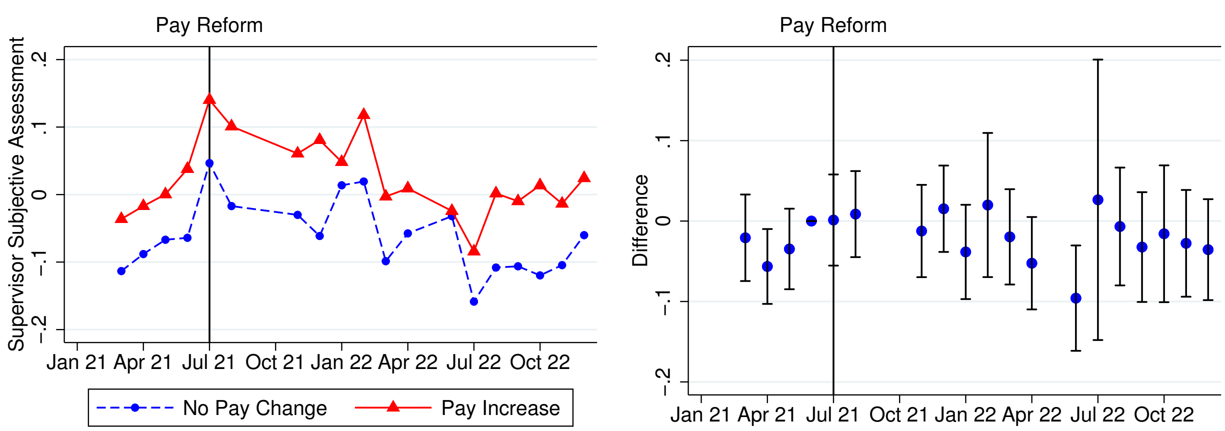
(c) Citizen Score

(d) Effort Metric

Note: This figure presents visualizations for the effect of the salary increase on secretary performance. We estimate Equation (2) and present dynamic DiD diagrams showing the coefficient for each month. The base month is June 2021, the month before the salary increase. In Sub-Figure (a), the dependent variable is the supervisor evaluation score, which is on a 0 to 100 scale. In Sub-Figure (b), the dependent variable is the auditor evaluation score, also on a 0 to 100 scale. In Sub-Figure (c), the dependent variable is the citizen evaluation score, which is a standardized index constructed from a Likert scale. In Sub-Figure (d), the dependent variable is the daily reporting rate, which is a proxy for the secretaries' effort. Table 1 shows the results from DiD regressions using these outcome variables. Results are discussed in Section 4.1.

Figure A4: Effect of Salary Increase on Subjective Performance Evaluation of GP Secretary

(a) Raw Time Trend (b) Dynamic Diff-in-Diff

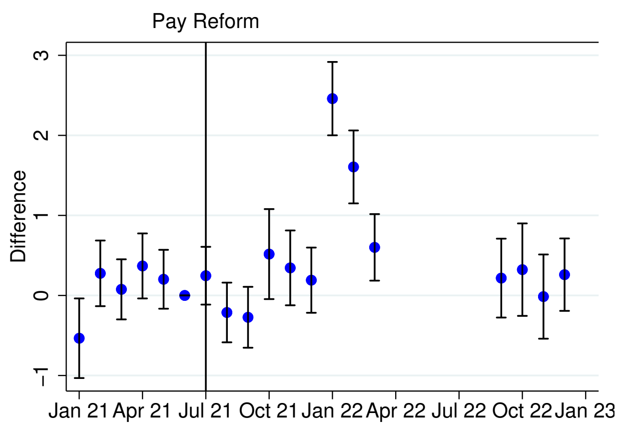


Note: This figure presents visualizations of the effect of the salary increase on subjective secretary performance. Sub-Figure (a) shows time trends of the subjective supervisor assessment using minimally processed data. Sub-Figure (b) presents dynamic DiD coefficient estimates for subjective supervisor assessment from estimating Equation (2). In Sub-figure (a), the outcomes are expressed as SDs from the mean. Table A4 shows the results from DiD regressions using these outcome variables. See discussion in Section 4.1. Results are discussed in Section 4.2.

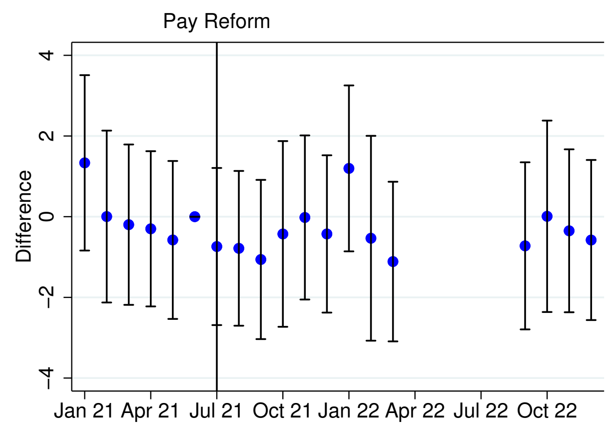
Figure A5: Effect of Salary Increase on Honesty and Corruption

Misreporting

(a) Self Reported Score

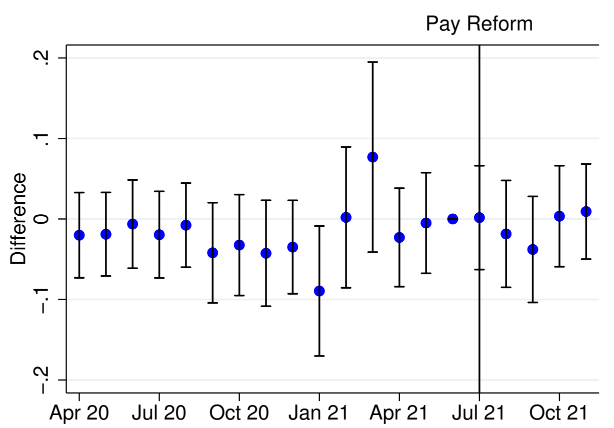


(b) Score Inflation

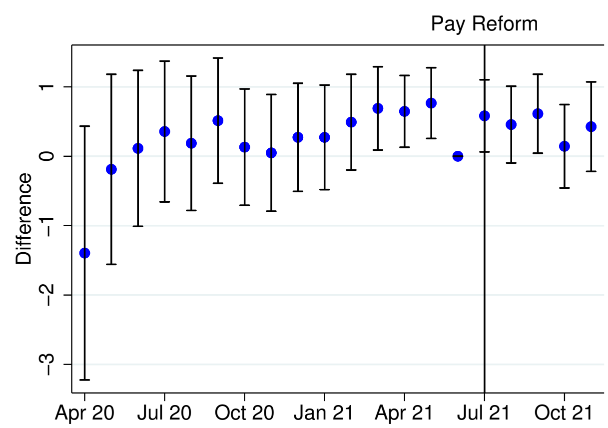


Misappropriations

(c) Secretary Self-Reimbursements

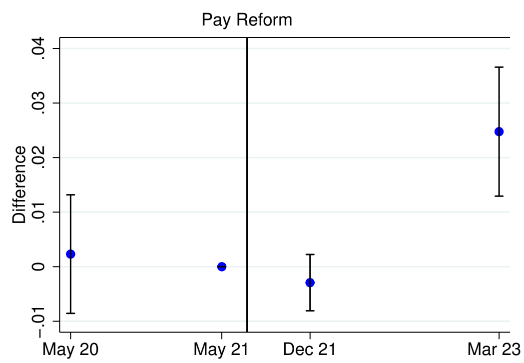


(d) Nepotism in Hiring



Note: We estimate Equation (2) using the honesty and corruption outcomes and present the dynamic DiD coefficients for each month. The base month is June 2021. Sub-Figure (a) shows the DID timelines for the secretaries' self-reported assessment. Sub-Figure (b) shows the DID timelines of secretaries' score inflation for overlapping components, as defined in Section 3.3.1. In Sub-Figure (c), the outcome variable is the share of GP expenditure that was spent reimbursing the GP secretary for personal expenses. In Sub-Figure (d), the outcome variable is the share of multipurpose workers employed by the GP who share a name with the GP Sarpanch. The vertical lines indicate July 2021, the month of the salary increase. Table 2 shows the results from DiD regressions using these outcome variables. Figure 2 shows the time trends for these outcomes using minimally processed data. Results are discussed Section 4.2.

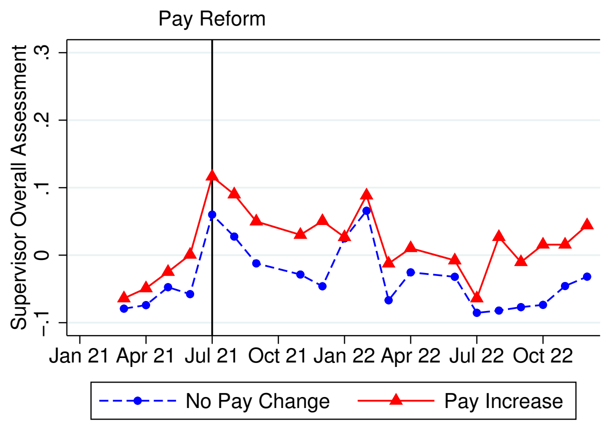
Figure A6: Effect of Salary Increase on GP Secretary Retention (Dynamic Diff-in-Diff)



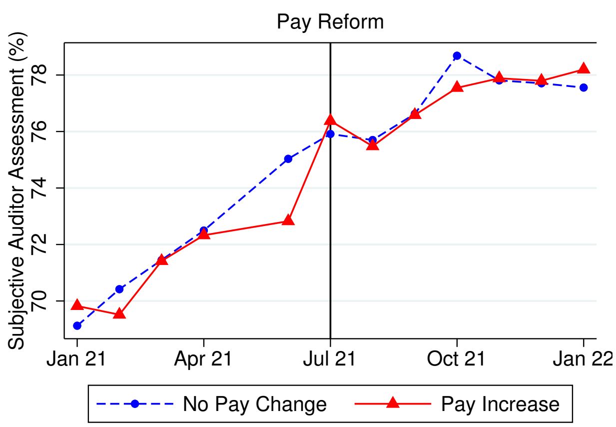
Notes: This figure presents DiD coefficients from estimating Equation (2). The base month is May 2021, the last pre-period. The dependent variable is a dummy indicating whether bureaucrat i is employed by the PRRD Department in period t. Sub-Figure (a) shows the DID coefficients of the share of GP secretaries holding a position in April 2020, continuing to hold a position within any of the GP or Mandal administrations. Table 3 shows the DiD estimate of Sub-figure (a). See discussion in Section 4.3.

Figure A7: Effect of Salary Increase on Subjective Performance Evaluation

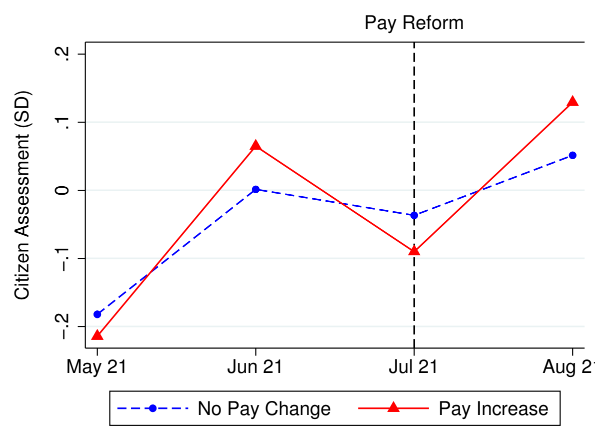
(a) Supervisor Subjective Evaluation



(b) Auditor Subjective Evaluation



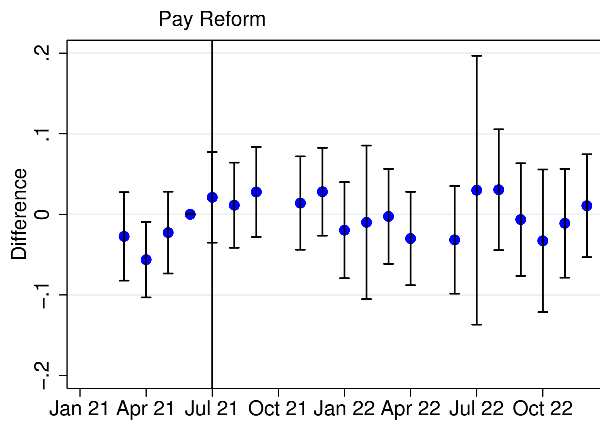
(c) Citizen Subjective Evaluation



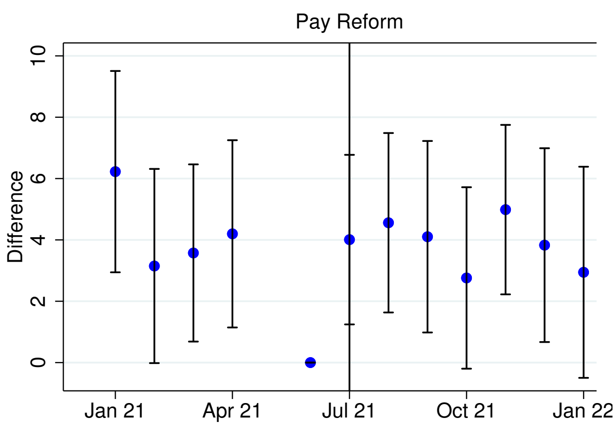
Note: This figure presents time trends of subjective performance evaluations using minimally processed data. Sub-Figure (a) shows time trends for the subjective performance evaluation by supervisors. In Sub-Figure (b), the dependent variable is subjective evaluations by state-level auditors. In Sub-Figure (c), the outcome variable is citizens' subjective evaluation of their GP secretary. The outcome in Sub-figure (b) is expressed as a percentage out of 100. In Sub-figure (a) and (c) the outcomes are expressed as SDs from the mean. Table A8 shows the DiD regression results using these outcome variables. See discussion in Section 4.4.

Figure A8: Effect of Salary Increase on Subjective Performance Evaluations

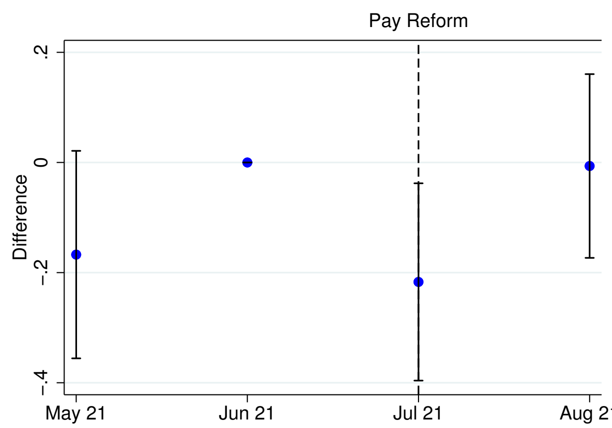
(a) Supervisor Subjective Assessment



(b) Auditor Subjective Evaluation

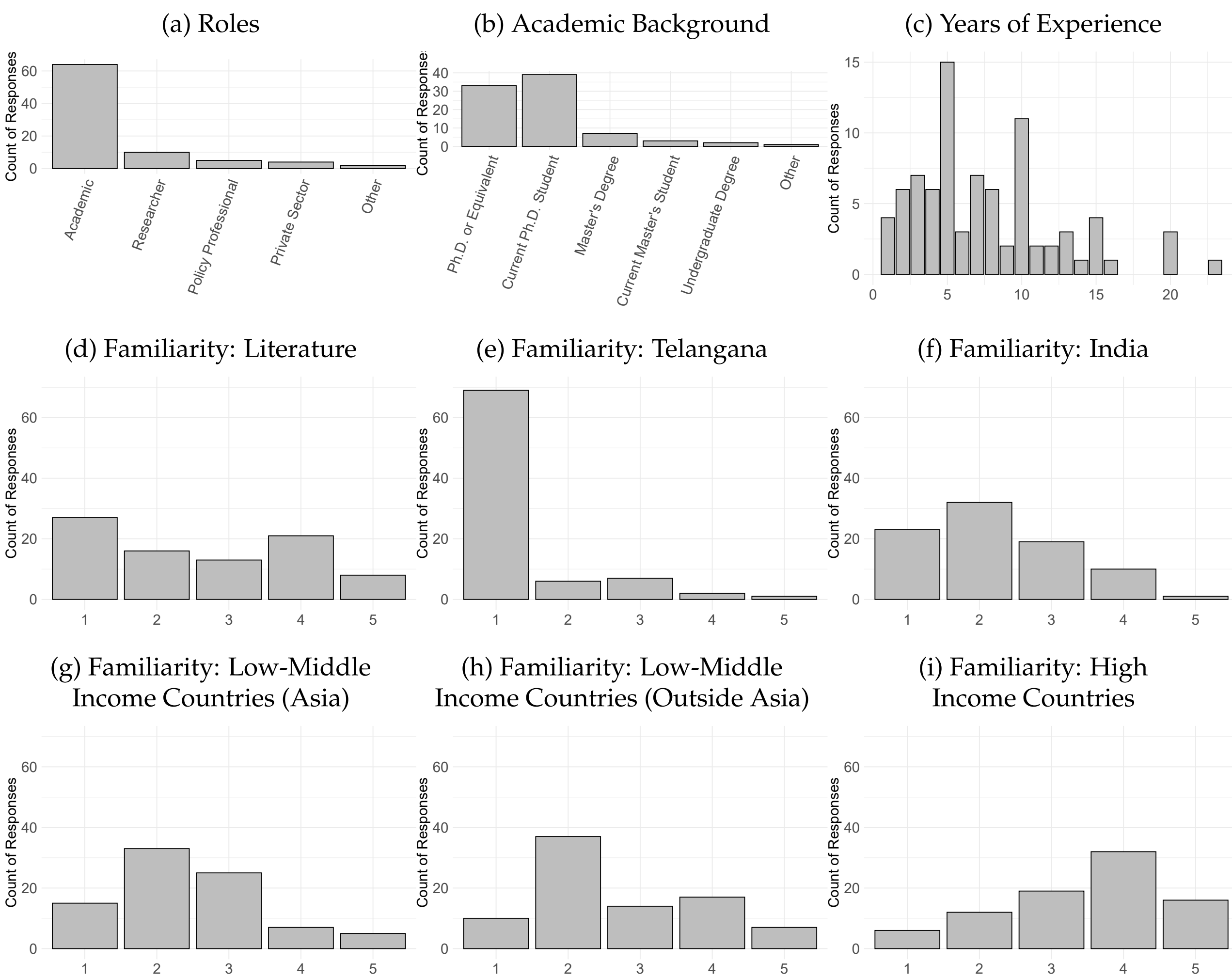


(c) Citizen Subjective Assessment



Note: This figure presents visualizations for the effect of the salary increase on subjective overall performance. We present DiD diagrams from estimating Equation (2). In Sub-Figure (a), the outcome variable is the supervisors' subjective evaluation of secretaries. In Sub-Figure (b), the dependent variable is state-level auditors' subjective evaluation. In Sub-Figure (c), the dependent variable is citizens' subjective evaluation. The outcome in Sub-figure (a) and (c) is expressed as as as SDs from the mean. In Sub-figure (b), the outcome is expressed as a percentage out of 100. Table A8 shows the results from DiD regressions using these outcome variables. See discussion in Section 4.4.

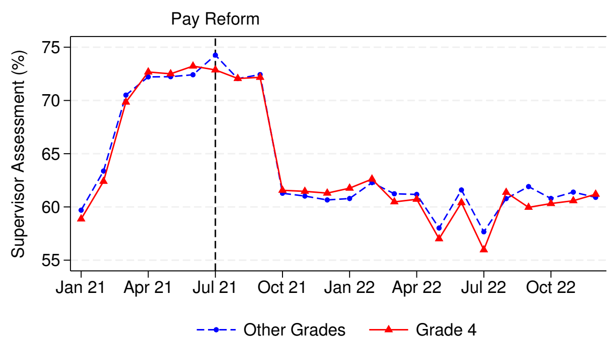
Figure A9: Forecaster Characteristics



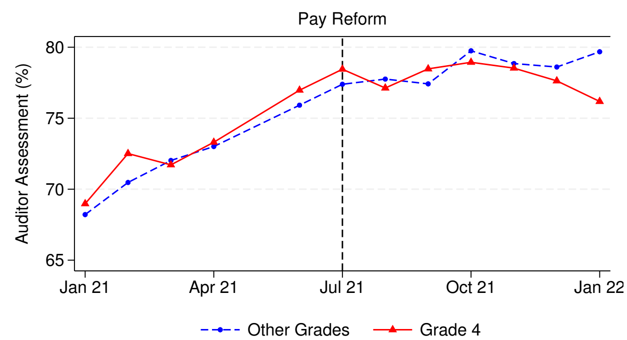
Note: This figure presents summary statistics on the characteristics of experts who made forecasts via the Social Science Prediction Platform. A total of 85 respondents completed our survey. Table 4 shows summary statistics on the forecasts. For subfigures (c) to (g), 1 to 5 corresponds to the following responses: "Not familiar at all", "Slightly Familiar", "Moderately Familiar", "Very Familiar" and "Extremely Familiar". Results are discussed in Section 4.5.

Figure A10: Relative Pay Effects: Comparing Performance of Grade IV and Grade I-III PSes

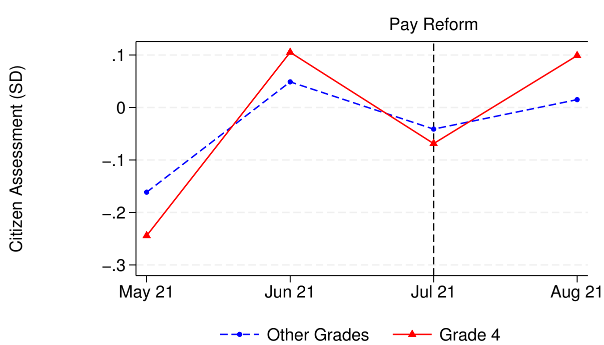
(a) Supervisor Assessment



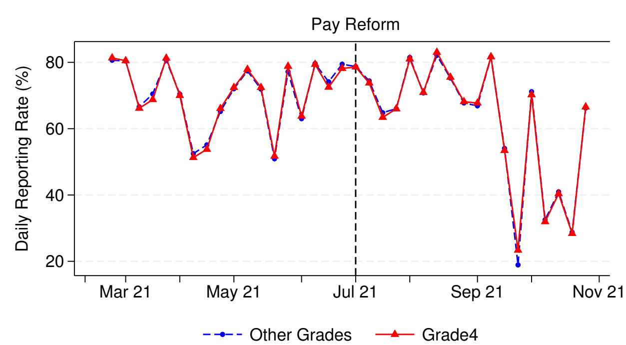
(b) Auditor Assessment



(c) Citizen Assessment

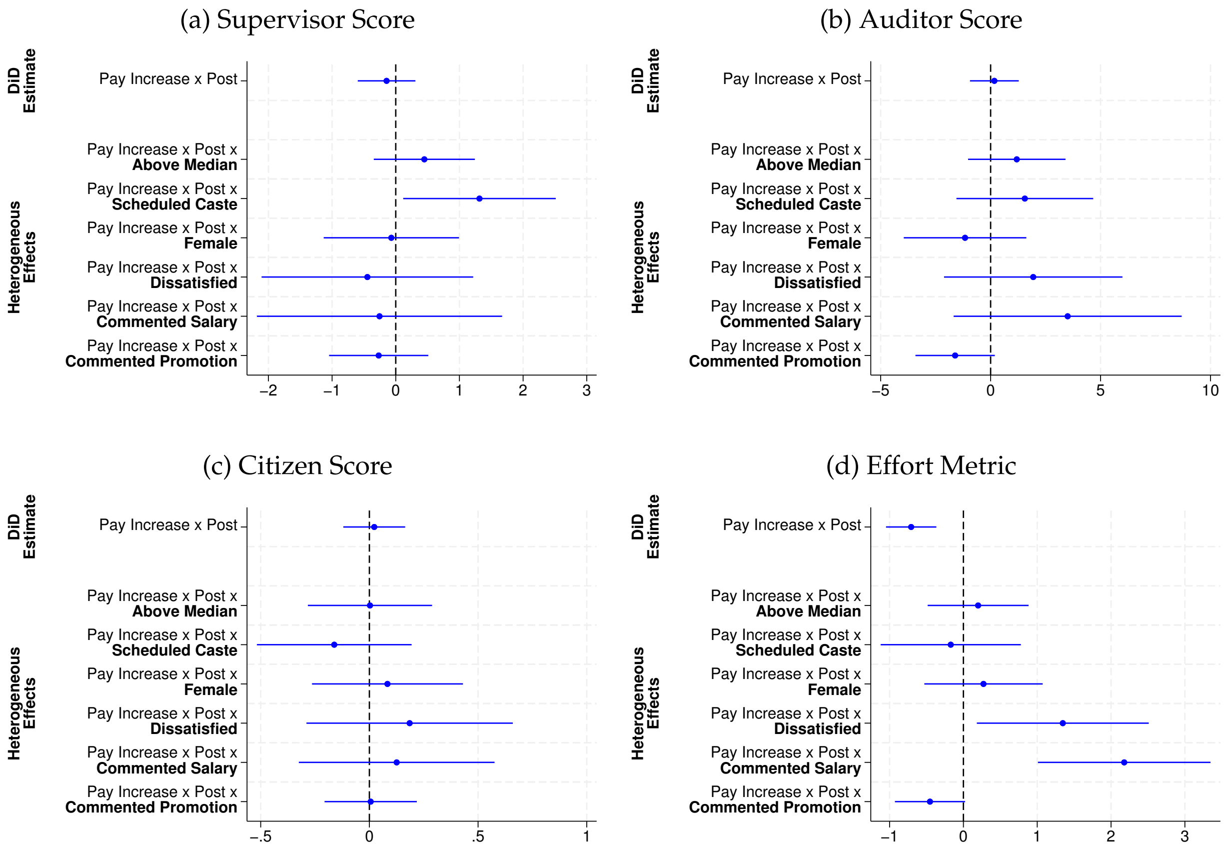


(d) Effort Metric



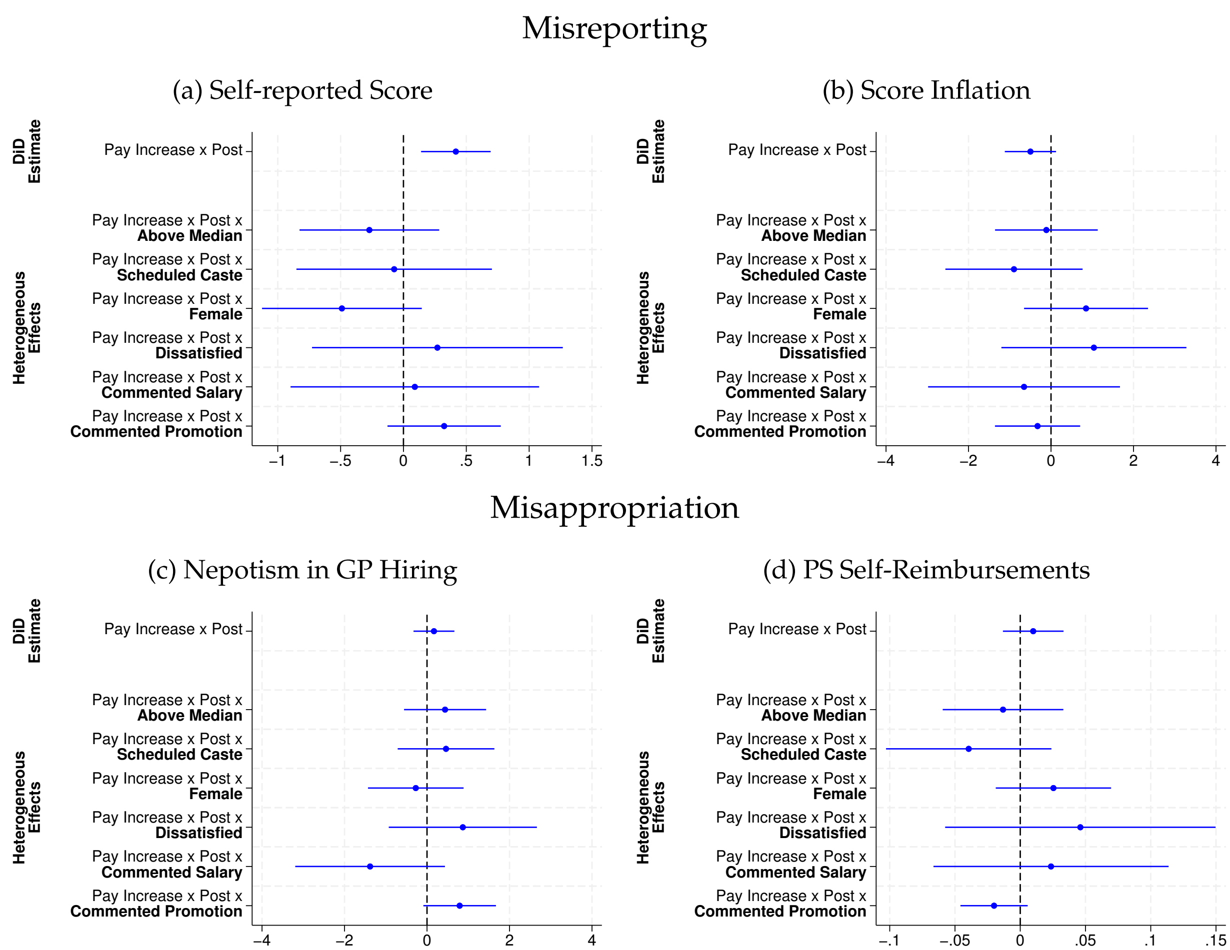
Note: This figure presents visualizations of the effect of pay rank. We present averages of minimally processed data for the four performance outcome variables. In each figure, we present monthly averages for GPs run by Grade IV PSes (whose salaries became equal to those of Junior PSes after the pay change) and GPs run by Regular PSes in higher salary grades (Grade I-III), whose pay rank was unaffected by the Junior PS pay hike. . Sub-Figure (a) shows timelines for the performance scores calculated from the supervisor assessments. Sub-Figure (b) shows the timelines for performance scores calculated from state-level auditors. Sub-Figure (c) shows the timelines for the performance scores calculated from the citizen assessments. Sub-Figure (d) shows how regularly the secretary files the required daily sanitation report, which we use as a proxy for effort. In both sub-figure (a) and (b) the outcomes are expressed as percentage out of 100. The outcome in Sub-figure (c) is expressed as SDs from the mean. The outcome in Sub-figure (d) is expressed as the percentage of total working days. Table A11 shows the results from DiD regressions using these outcome variables. See discussion in Section 5.3.

Figure A11: Effect of Salary Increase on Performance — Heterogeneity



Note: We present coefficient plots of the DiD estimates from Equation (1). The outcome variables are the performance measures based on supervisor, auditor and citizen evaluations and the effort proxy. In each Sub-Figure, the first coefficient is the DiD Estimate, which presents the base DiD estimate. The *Heterogeneous Effects* coefficients present the coefficient on the triple interaction between *Pay Increase*, *Post* and the secretary characteristics highlighted in bold. For the characteristic *Commented Promotion*, the interaction term *Post* × *Commented Promotion* is excluded due to collinearity. Each coefficient comes from a different regression (i.e, we run one regression for each dimension of heterogeneity). These regression coefficients are also presented in tables A13 and A14. Refer to the discussion in Section 5.

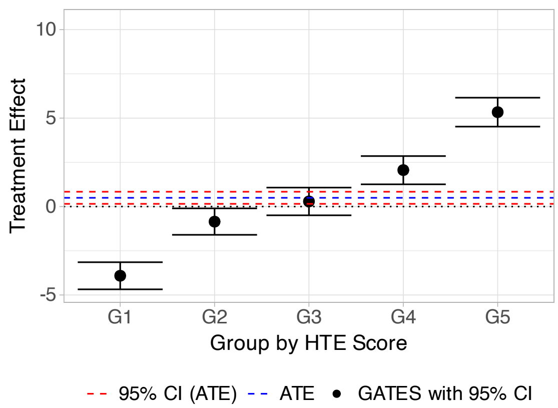
Figure A12: Heterogeneity in the Effect on Honesty and Corruption



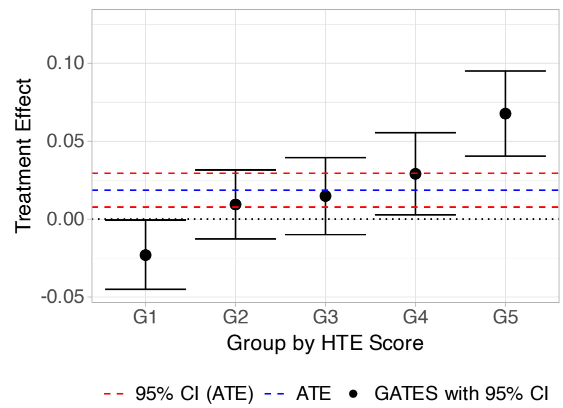
Note: We present coefficient plots of the DiD estimates from Equation (1). The outcome variables in Sub-Figures (a)-(b) are measures of honesty: self-reported performance score in Sub-Figure (a) and score inflation in Sub-Figure (b). The outcome variables in Sub-Figures (c)-(d) are proxies of corruption: the share of GP multi-purpose workers who share the Sarpanch's name in (c) and the share of GP expenditure reimbursed to the secretary for personal expenses in (d). In each Sub-Figure, the first coefficient is the DiD Estimate, which presents the base DiD estimate. The *Heterogeneous Effects* coefficients present the coefficient on the triple interaction between *Pay Increase*, *Post* and the secretary characteristics highlighted in bold. Each coefficient comes from a different regression (i.e, we run one regression for each dimension of heterogeneity). For the characteristic *Commented Promotion*, the interaction term *Post* × *Commented Promotion* is excluded due to collinearity. These regression coefficients are also presented in tables A15 and A16. Refer to the discussion in Section 5.

Figure A13: Generic ML Predicted Treatment Effects on Supervisor Evaluation, Retention and Effort

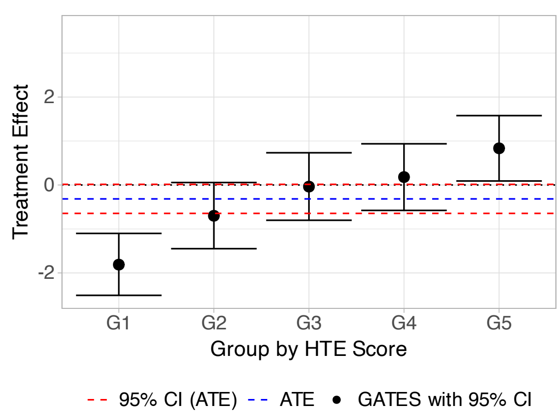
(a) GATES: Supervisor Assessment



(b) GATES: Stay Rate



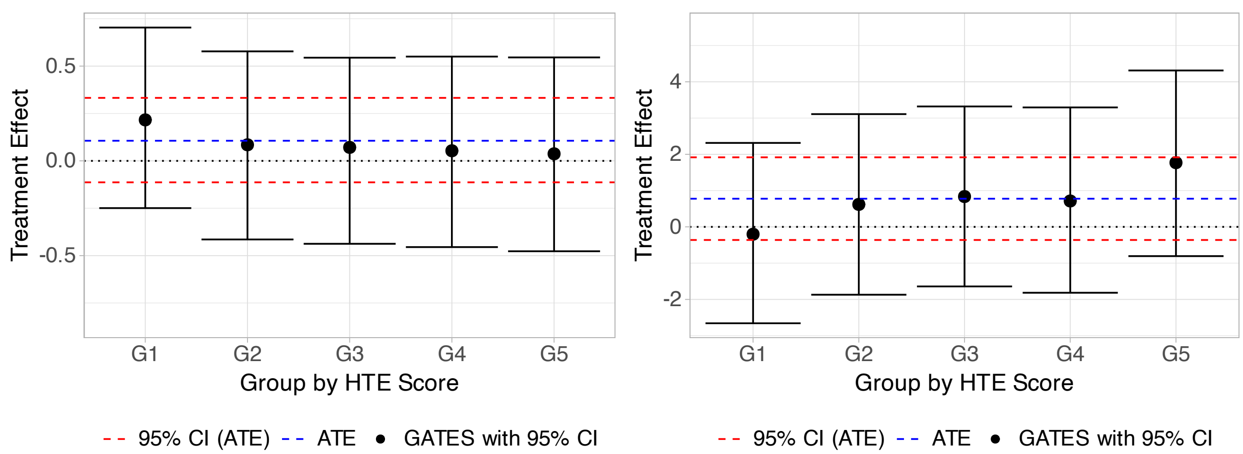
(c) GATES: Effort Metric



Note: This figure displays Sorted Grouped Average Treatment Effects (GATES), as described in Section B. We use the GenericML model to predict treatment effects for each secretary and sort secretaries into 5 groups (quintiles) based on their predicted treatment effect for that outcome. We present the GATES generated by GenericML for 3 variables: supervisor score, secretary stay rate and the daily reporting rate, our measure of bureaucrat effort. The x-axis represents the quintile of the predicted treatment effect. Results are discussed in Section 5.

Figure A14: Predicted Treatment Effects — Citizen and Auditor Evaluations

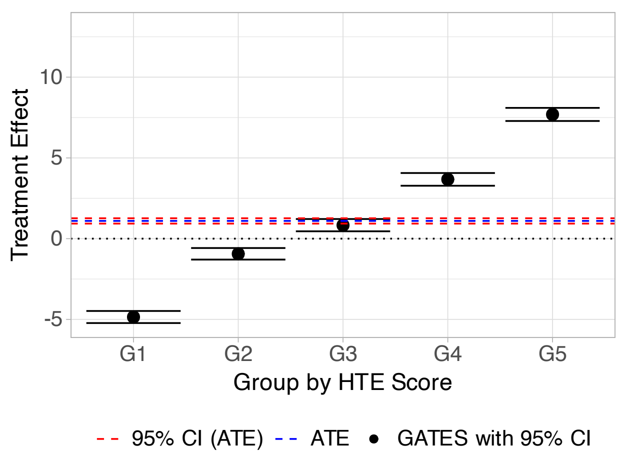
(a) GATES: Citizen Assessment (b) GATES: Auditor Assessment



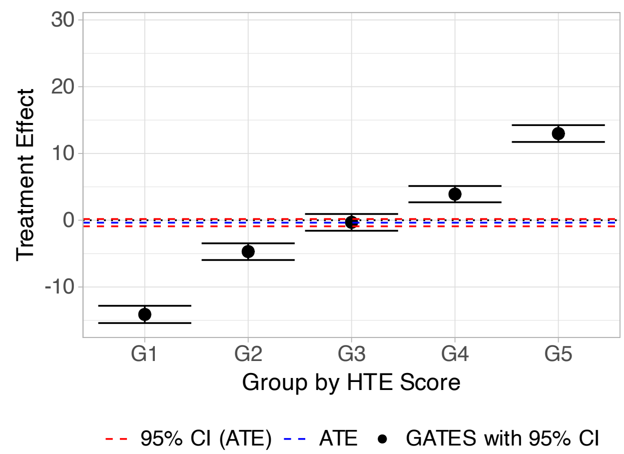
Note: Note: This figure displays Sorted Grouped Average Treatment Effects (GATES), as described in Section B. We use the GenericML model to predict treatment effects for each secretary and sort secretaries into 5 groups (quintiles) based on their predicted treatment effect for that outcome. We present the GATES generated by GenericML for two outcomes: auditor evaluations and citizen evaluations. The x-axis represents the quintile of the predicted treatment effect. Results are discussed in Section 5.

Figure A15: Predicted Treatment Effects — Honesty and Corruption Outcomes

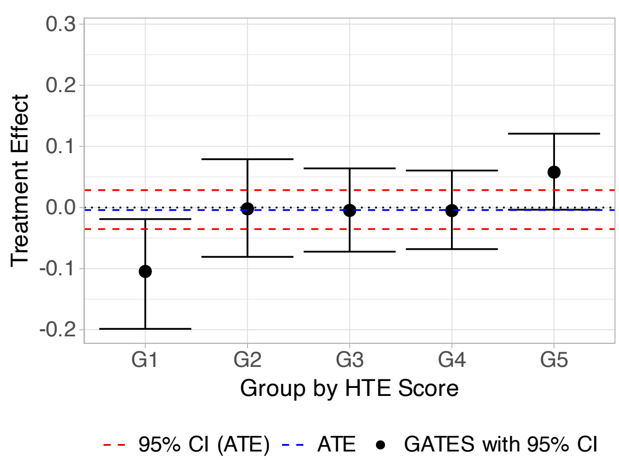
(a) GATES: PS Self-Assessed Score



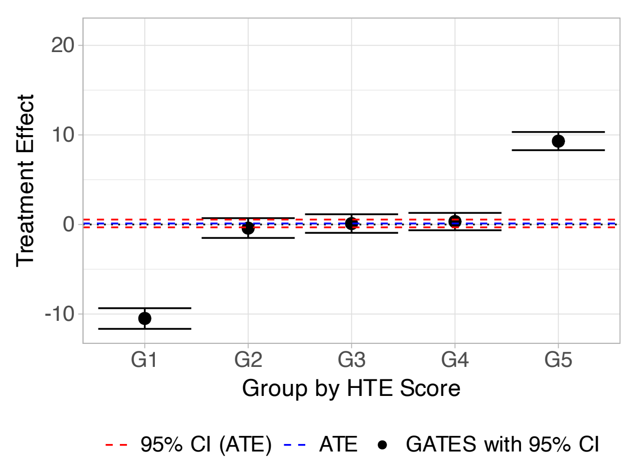
(b) GATES: Score Inflation



(c) GATES: PS Self-Reimbursements

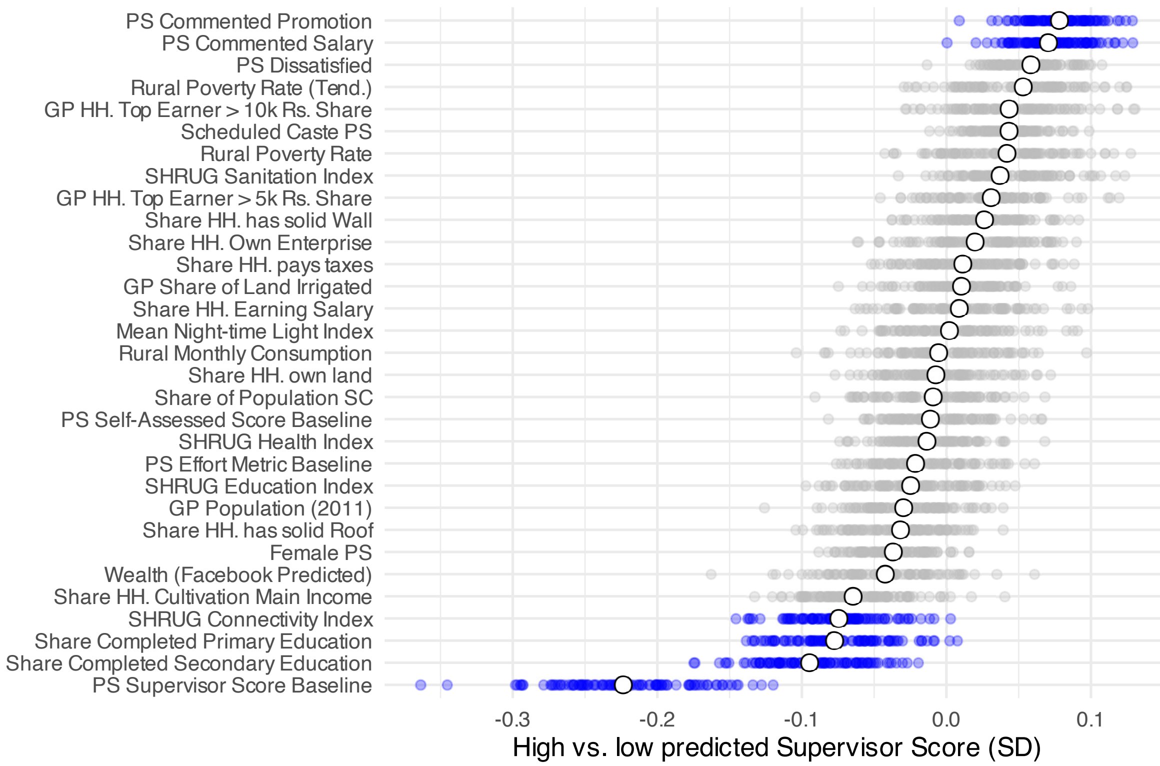


(d) GATES: Nepotism in Hiring



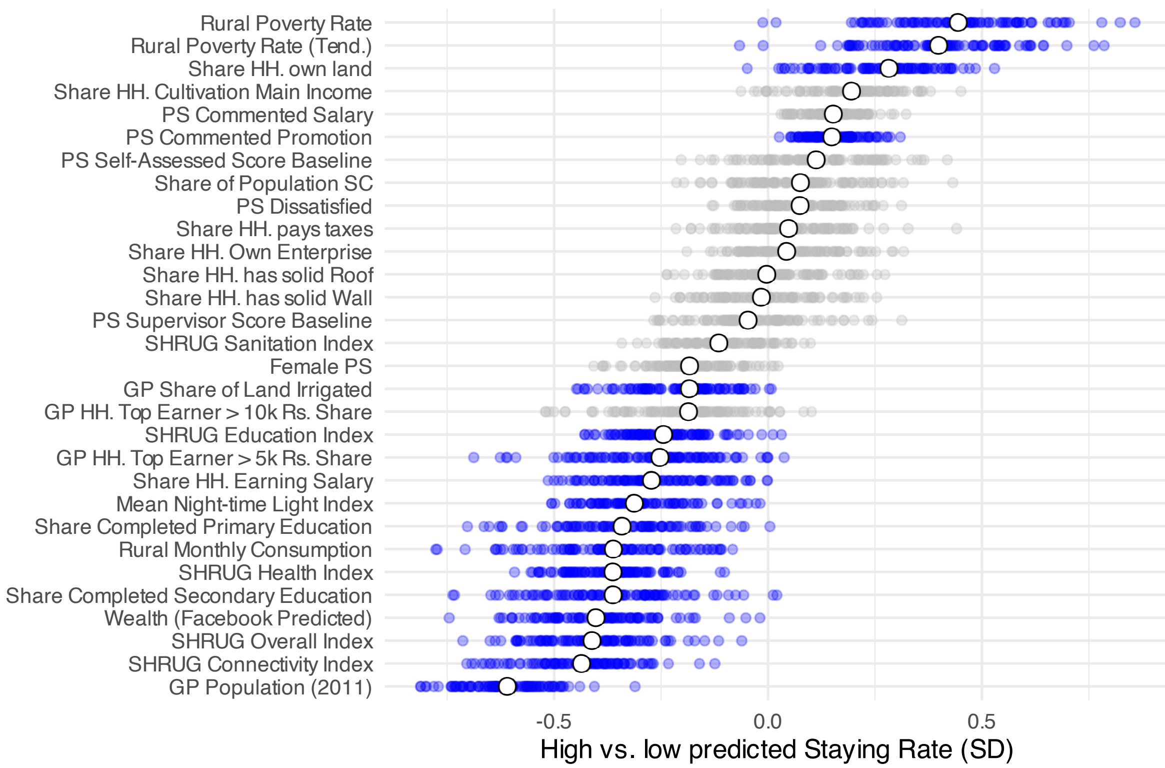
Note: This figure displays Sorted Grouped Average Treatment Effects (GATES), as described in Section B. We use the GenericML model to predict treatment effects for each secretary and sort secretaries into 5 groups (quintiles) based on their predicted treatment effect for that outcome. We present the GATES generated by GenericML for four variables: self-reported performance, score inflation, share of expenditure reimbursed to the secretary for personal expenses, and the share of multipurpose workers employed by the GP who share a name with the Sarpanch. The x-axis represents the quintile of predicted the treatment effect. Results are discussed in Section 5.

Figure A16: GenericML Classification Analysis — Supervisor Evaluation



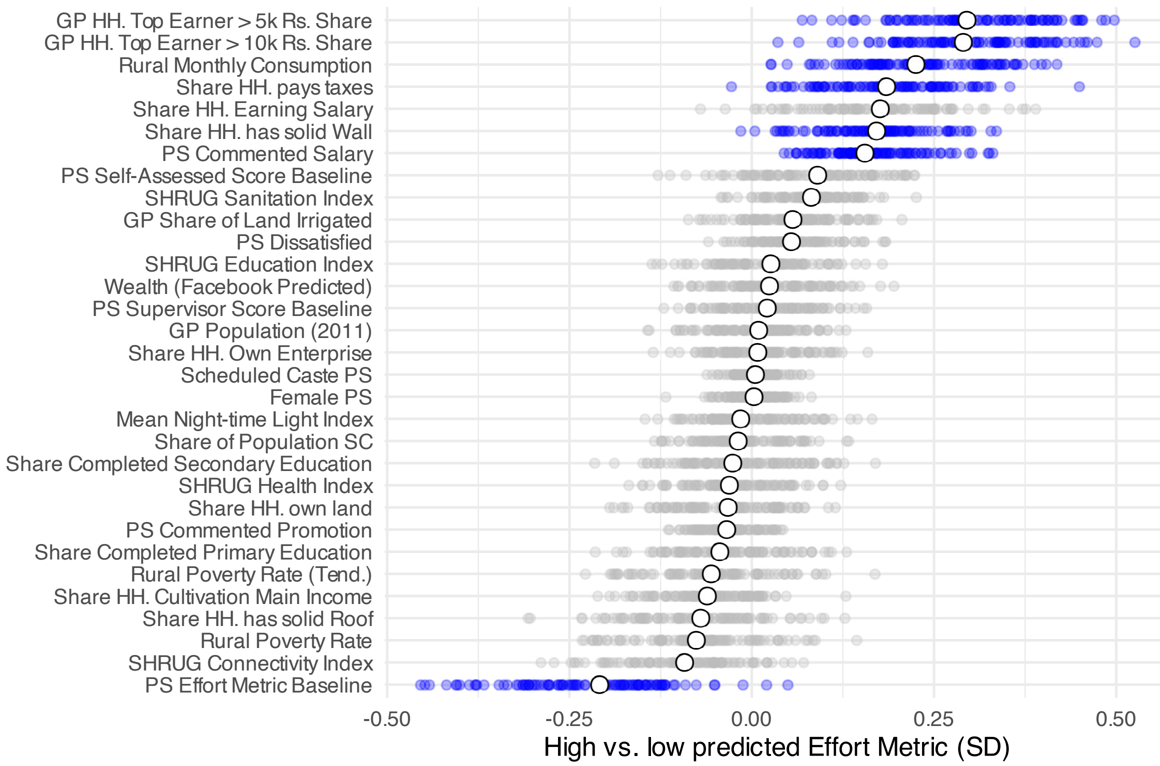
Notes: This figure presents Classification Analysis (CLAN) estimates from the GenericML model, as described in Section B. The outcome is supervisor evaluation scores. Each row is a baseline secretary or GP characteristic that is used as a predictor in the GenericML model. For each characteristic, we show the mean difference between observations (i.e. secretaries/GPs) predicted to be in the top quintile of treatment effects and observations (i.e. secretaries/GPs) predicted to be in the bottom quintile of treatment effects. We standardise each variable so readers can easily interpret magnitudes. For each sample split (i.e. iteration of the GenericML estimation), we obtain different predicted treatment effects and a different high vs low mean difference. Each dot plots one split for a given covariate. Large circles with white centers plot the median difference across all sample splits. Covariates for which the mean difference is statistically significant (at the 5% level) in ≥90% of splits are coloured blue. Results are discussed in Section 5.

Figure A17: GenericML Classification Analysis — Stay Rates



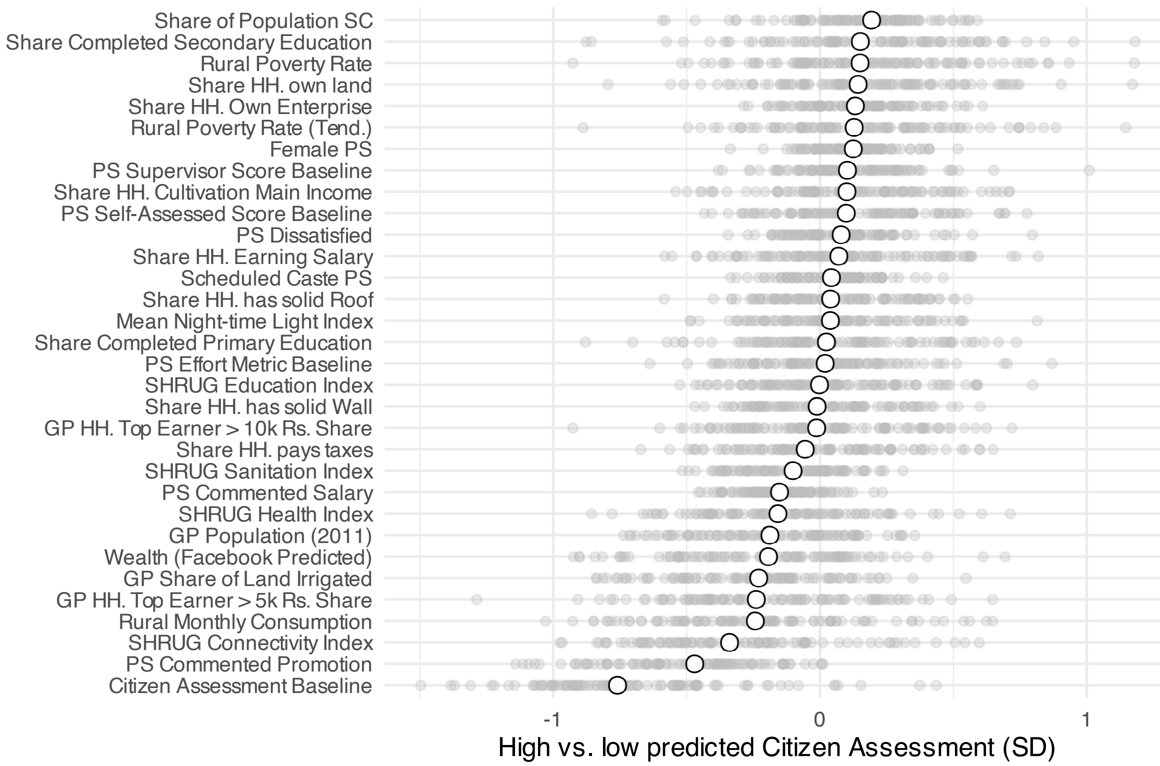
Notes: This figure presents Classification Analysis (CLAN) estimates from the GenericML model, as described in Section B. The outcome is GP secretary stay rates. Each row is a baseline secretary or GP characteristic that is used as a predictor in the GenericML model. For each characteristic, we show the mean difference between observations (i.e. secretaries/GPs) predicted to be in the top quintile of treatment effects and observations (i.e. secretaries/GPs) predicted to be in the bottom quintile of treatment effects. We standardise each variable so readers can easily interpret magnitudes. For each sample split (i.e. iteration of the GenericML estimation), we obtain different predicted treatment effects and a different high vs low mean difference. Each dot plots one split for a given covariate. Large circles with white centers plot the median difference across all sample splits. Covariates for which the mean difference is statistically significant (at the 5% level) in ≥90% of splits are coloured blue. Results are discussed in Section 5.

Figure A18: GenericML Classification Analysis — Effort



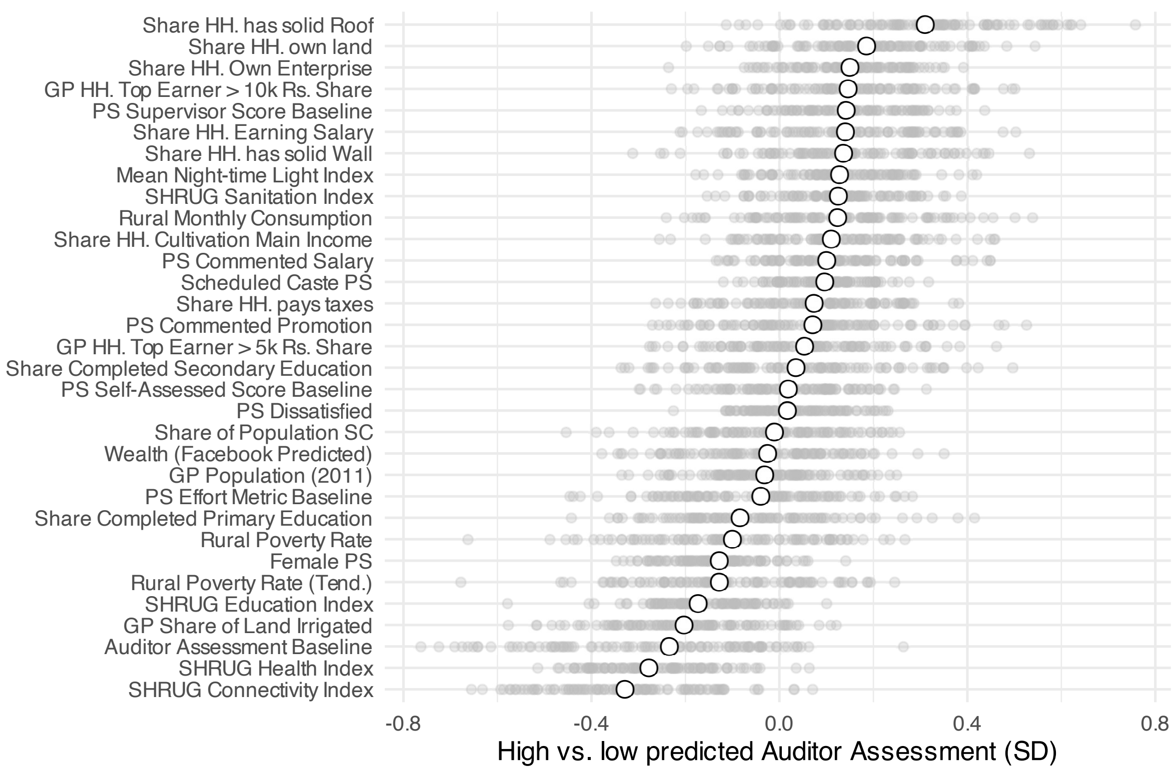
Notes: This figure presents Classification Analysis (CLAN) estimates from the GenericML model, as described in Section B. The outcome is the daily reporting rate, our measure of bureaucrat effort. Each row is a baseline secretary or GP characteristic that is used as a predictor in the GenericML model. For each characteristic, we show the mean difference between observations (i.e. secretaries/GPs) predicted to be in the top quintile of treatment effects and observations (i.e. secretaries/GPs) predicted to be in the bottom quintile of treatment effects. We standardise each variable so readers can easily interpret magnitudes. For each sample split (i.e. iteration of the GenericML estimation), we obtain different predicted treatment effects and a different high vs low mean difference. Each dot plots one split for a given covariate. Large circles with white centers plot the median difference across all sample splits. Covariates for which the mean difference is statistically significant (at the 5% level) in $\geq$90% of are coloured blue. Results are discussed in Section 5.

Figure A19: GenericML Classification Analysis — Citizen Evaluations



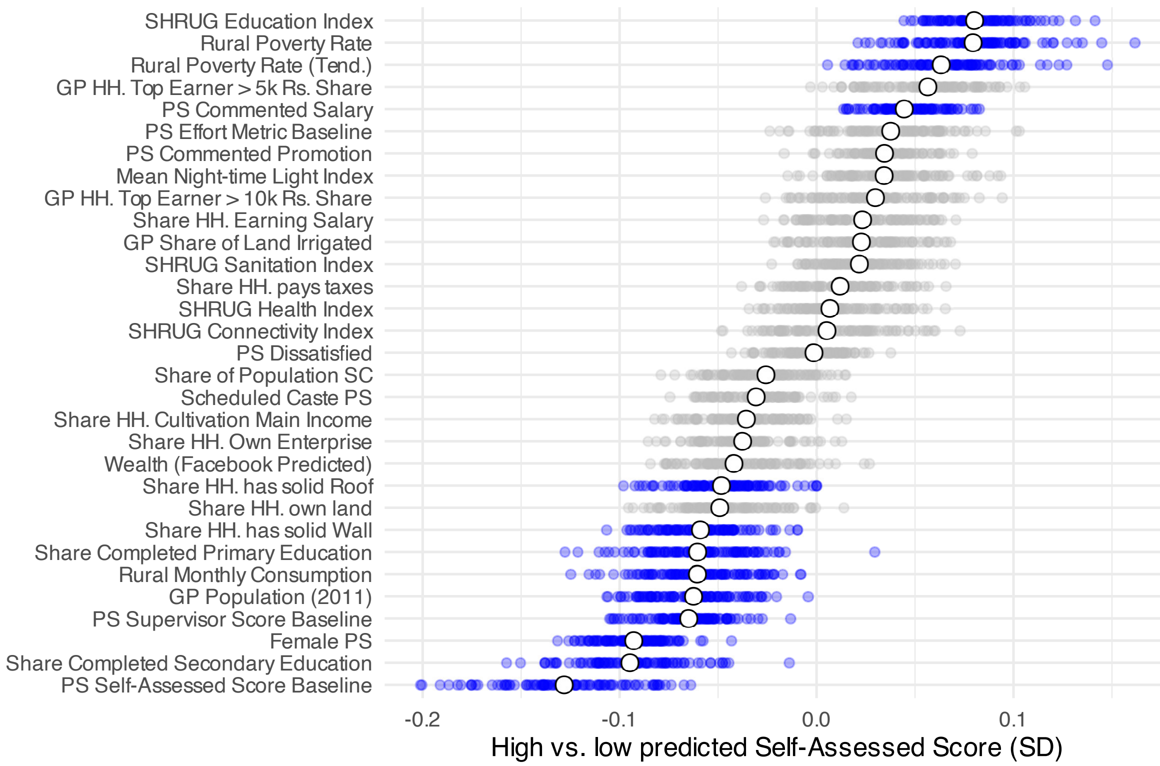
Notes: This figure presents Classification Analysis (CLAN) estimates from the GenericML model, as described in Section B. The outcome is citizen evaluation scores. Each row is a baseline secretary or GP characteristic that is used as a predictor in the GenericML model. For each characteristic, we show the mean difference between observations (i.e. secretaries/GPs) predicted to be in the top quintile of treatment effects and observations (i.e. secretaries/GPs) predicted to be in the bottom quintile of treatment effects. We standardise each variable so readers can easily interpret magnitudes. For each sample split (i.e. iteration of the GenericML estimation), we obtain different predicted treatment effects and a different high vs low mean difference. Each dot plots one split for a given covariate. Large circles with white centers plot the median difference across all sample splits. Covariates for which the mean difference is statistically significant (at the 5% level) in ≥90% of splits are coloured blue. Results are discussed in Section 5.

Figure A20: GenericML Classification Analysis — Auditor Evaluations



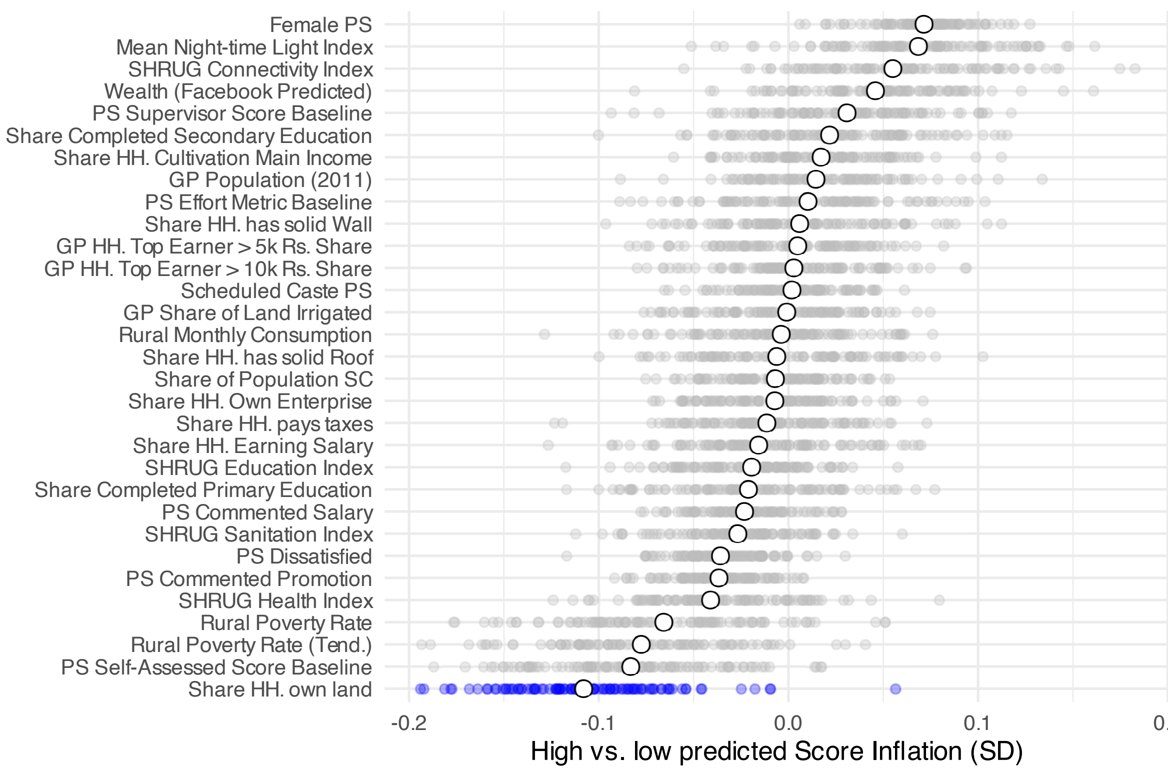
Notes: This figure presents Classification Analysis (CLAN) estimates from the GenericML model, as described in Section B. The outcome is auditor evaluation scores. Each row is a baseline secretary or GP characteristic that is used as a predictor in the GenericML model. For each characteristic, we show the mean difference between observations (i.e. secretaries/GPs) predicted to be in the top quintile of treatment effects and observations (i.e. secretaries/GPs) predicted to be in the bottom quintile of treatment effects. We standardise each variable so readers can easily interpret magnitudes. For each sample split (i.e. iteration of the GenericML estimation), we obtain different predicted treatment effects and a different high vs low mean difference. Each dot plots one split for a given covariate. Large circles with white centers plot the median difference across all sample splits. Covariates for which the mean difference is statistically significant (at the 5% level) in ≥90% of splits are coloured blue. Results are discussed in Section 5.

Figure A21: GenericML Classification Analysis — Self-Reported Score



Notes: This figure presents Classification Analysis (CLAN) estimates from the GenericML model, as described in Section B. The outcome is self-assessed performance scores. Each row is a baseline secretary or GP characteristic that is used as a predictor in the GenericML model. For each characteristic, we show the mean difference between observations (i.e. secretaries/GPs) predicted to be in the top quintile of treatment effects and observations (i.e. secretaries/GPs) predicted to be in the bottom quintile of treatment effects. We standardise each variable so readers can easily interpret magnitudes. For each sample split (i.e. iteration of the GenericML estimation), we obtain different predicted treatment effects and a different high vs low mean difference. Each dot plots one split for a given covariate. Large circles with white centers plot the median difference across all sample splits. Covariates for which the mean difference is statistically significant (at the 5% level) in ≥90% of splits are coloured blue. Results are discussed in Section 5.

Figure A22: GenericML Classification Analysis — Score Inflation



Notes: This figure presents Classification Analysis (CLAN) estimates from the GenericML model, as described in Section B. The outcome is score inflation. Each row is a baseline secretary or GP characteristic that is used as a predictor in the GenericML model. For each characteristic, we show the mean difference between observations (i.e. secretaries/GPs) predicted to be in the top quintile of treatment effects and observations (i.e. secretaries/GPs) predicted to be in the bottom quintile of treatment effects. We standardise each variable so readers can easily interpret magnitudes. For each sample split (i.e. iteration of the GenericML estimation), we obtain different predicted treatment effects and a different high vs low mean difference. Each dot plots one split for a given covariate. Large circles with white centers plot the median difference across all sample splits. Covariates for which the mean difference is statistically significant (at the 5% level) in ≥90% of splits are coloured blue. Results are discussed in Section 5.

Figure A23: GenericML Classification Analysis — Secretary Self-Reimbursement

Notes: This figure presents Classification Analysis (CLAN) estimates from the GenericML model, as described in Section B. The outcome is the share of GP expenditure on reimbursing the secretary for personal expenses, a proxy for corruption. Each row is a baseline secretary or GP characteristic that is used as a predictor in the GenericML model. For each characteristic, we show the mean difference between observations (i.e. secretaries/GPs) predicted to be in the top quintile of treatment effects and observations (i.e. secretaries/GPs) predicted to be in the bottom quintile of treatment effects. We standardise each variable so readers can easily interpret magnitudes. For each sample split (i.e. iteration of the GenericML estimation), we obtain different predicted treatment effects and a different high vs low mean difference. Each dot plots one split for a given covariate. Large circles with white centers plot the median difference across all sample splits. Covariates for which the mean difference is statistically significant (at the 5% level) in ≥90% of splits are coloured blue. Results are discussed in Section 5.

Figure A24: GenericML Classification Analysis — Nepotism in Hiring

Notes: This figure presents Classification Analysis (CLAN) estimates from the GenericML model, as described in Section B. The outcome is share of multipurpose workers hired by the GP who share a name with the Sarpanch, a proxy for nepotism in hiring and corruption. Each row is a baseline secretary or GP characteristic that is used as a predictor in the GenericML model. For each characteristic, we show the mean difference between observations (i.e. secretaries/GPs) predicted to be in the top quintile of treatment effects and observations (i.e. secretaries/GPs) predicted to be in the bottom quintile of treatment effects. We standardise each variable so readers can easily interpret magnitudes. For each sample split (i.e. iteration of the GenericML estimation), we obtain different predicted treatment effects and a different high vs low mean difference. Each dot plots one split for a given covariate. Large circles with white centers plot the median difference across all sample splits. Covariates for which the mean difference is statistically significant (at the 5% level) in ≥90% of splits are colored blue. Results are discussed in Section 5.

D Additional Tables

Table A1: Summary Statistics — Baseline Characteristics of GP Secretaries

	Pay Increase		No Pay Change		
	Obs. (Cluster)	Mean (SD)	Obs. (Cluster)	Mean (SD)	Diff. (SE)
Panel A. Baseline Performance					
Supervisor Assessment (%)	8,976	67.69	3,645	67.51	0.179
	(8,976)	(11.58)	(3,645)	(11.93)	(0.23)
Supervisor Overall GP Assmt. (SD)	8,496	0.01	3,476	-0.02	0.028
	(8,496)	(1.00)	(3,476)	(1.00)	(0.02)
Supervisor Subjective Secretary Assmt. (SD)	8,499	0.02	3,478	-0.06	0.080***
	(8,499)	(1.00)	(3,478)	(1.00)	(0.02)
Self-reported Assessment (%)	9,081	69.04	3,689	67.99	1.052***
	(9,081)	(7.10)	(3,689)	(7.18)	(0.14)
Citizen Assessment (SD)	1,707	-0.06	688	-0.05	-0.013
	(1,707)	(0.86)	(688)	(0.86)	(0.04)
Citizen Overall Assessment (SD)	1,678	-0.04	680	-0.11	0.066
	(1,678)	(0.88)	(680)	(0.89)	(0.04)
Reporting Rate (%)	9,081	70.95	3,689	70.31	0.639***
	(9,081)	(8.80)	(3,689)	(10.14)	(0.19)
Auditor Assessment (%)	4,263	71.98	1,723	71.77	0.207
	(4,263)	(12.42)	(1,723)	(13.34)	(0.37)
Auditor Overall Assessment (%)	4,262	71.07	1,723	71.27	-0.205
	(4,262)	(13.83)	(1,723)	(13.89)	(0.40)
Sarpanch Self-Dealing (%)	7,125	9.26	2,727	6.20	3.058***
	(7,125)	(23.04)	(2,727)	(18.66)	(0.45)
Secretary Self-Reimbursements (%)	9,080	0.09	3,689	0.09	0.001
	(9,080)	(0.82)	(3,689)	(0.82)	(0.02)
Panel B. PS Demographics					
Female (%)	8,785	31.34	3,393	23.73	7.612***
	(8,785)	(46.39)	(3,393)	(42.55)	(0.88)
Age	8,100	31.15	2,692	40.98	-9.833***
	(8,100)	(4.30)	(2,692)	(7.23)	(0.15)
Years in Service	8,099	1.95	2,680	11.81	-9.864***
	(8,099)	(0.45)	(2,680)	(6.97)	(0.13)
Scheduled Castes (%)	9,081	17.85	3,688	16.08	1.771**
	(9,081)	(38.30)	(3,688)	(36.74)	(0.73)
Scheduled Tribes (%)	9,081	12.43	3,688	7.16	5.274***
	(9,081)	(33.00)	(3,688)	(25.78)	(0.55)
Panel C. PS Satisfaction Survey					
Dissatisfied (%)	7,114	15.36	2,360	12.12	3.25***
	(7,114)	(36.06)	(2,360)	(32.64)	(0.80)
Commented Salary (%)	7,114	20.23	2,360	8.69	11.54***
	(7,114)	(40.17)	(2,360)	(28.17)	(0.75)
Commented Promotion (%)	7,114	13.90	2,360	2.67	11.23***
	(7,114)	(34.60)	(2,360)	(16.12)	(0.53)

Notes: This table shows baseline summary statistics for treated secretaries (who received the pay increase) and comparison secretaries (who did not). Panel A shows the means of the performance indicators. Panel B shows summary statistics for the secretaries' demographics. Panel C shows average responses to questions about job satisfaction in the secretary survey. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 2.3.

Table A2: Summary Statistics — Baseline GP Characteristics

	Pay Increase		No Pay Change		
	Obs. (Cluster)	Mean (SD)	Obs. (Cluster)	Mean (SD)	Diff. (SE)
Panel A. GP Statistics					
Total Population (2011)	9,075	1,326.56	3,685	2,270.40	-943.835***
	(9,075)	(942.85)	(3,685)	(2,028.94)	(34.86)
Share Completed Secondary Education (2012)	8,812	0.18	3,552	0.19	-0.008***
	(8,812)	(0.08)	(3,552)	(0.08)	(0.00)
Share Completed Primary Education (2012)	8,812	0.45	3,552	0.46	-0.009***
	(8,812)	(0.11)	(3,552)	(0.11)	(0.00)
Share of Scheduled Caste (2012)	8,812	0.05	3,552	0.06	-0.002**
	(8,812)	(0.04)	(3,552)	(0.05)	(0.00)
Rural Monthly Consumption per capita (2011-12)	8,792	22,346.89	3,541	22,396.52	-49.626
	(8,792)	(4,308.54)	(3,541)	(4,303.53)	(85.68)
Consumption <31 INR per day per capita (2011-12)	8,792	0.11	3,541	0.11	-0.000
	(8,792)	(0.08)	(3,541)	(0.08)	(0.00)
Mean Night-time Light Index (2020)	8,885	0.75	3,603	0.89	-0.145***
	(8,885)	(0.77)	(3,603)	(1.20)	(0.02)
Wealth Index (Facebook Predicted) 2021	8,881	-0.16	3,601	-0.12	-0.046***
	(8,881)	(0.24)	(3,601)	(0.27)	(0.01)
Panel B. 2012 Household Statistics (Share of Households)					
Highest Earner >= 5k Rs	8,792	0.21	3,541	0.21	-0.000
	(8,792)	(0.19)	(3,541)	(0.19)	(0.00)
Share of total land irrigated	8,737	0.50	3,517	0.49	0.005
	(8,737)	(0.17)	(3,517)	(0.17)	(0.00)
Earning any Salaried Wage	8,792	0.05	3,541	0.05	-0.005***
	(8,792)	(0.08)	(3,541)	(0.09)	(0.00)
Owning Enterprises	8,792	0.03	3,541	0.03	0.000
	(8,792)	(0.09)	(3,541)	(0.09)	(0.00)
Main Income is Cultivation	8,792	0.32	3,541	0.31	0.004
	(8,792)	(0.23)	(3,541)	(0.23)	(0.00)
House has solid wall	8,812	0.89	3,552	0.88	0.004
	(8,812)	(0.16)	(3,552)	(0.16)	(0.00)
House has solid roof	8,812	0.71	3,552	0.70	0.008
	(8,812)	(0.29)	(3,552)	(0.28)	(0.01)
Pays Income or Professional Taxes	8,792	0.05	3,541	0.05	0.000
	(8,792)	(0.12)	(3,541)	(0.12)	(0.00)
Owning Land	8,792	0.51	3,541	0.50	0.015***
	(8,792)	(0.21)	(3,541)	(0.21)	(0.00)
Panel C. 2011 Census Indices					
Education Index	8,801	-0.01	3,547	0.03	-0.04***
	(8,801)	(0.37)	(3,547)	(0.39)	(0.01)
Health Index	8,801	-0.02	3,547	0.04	-0.06***
	(8,801)	(0.35)	(3,547)	(0.43)	(0.01)
Sanitation Index	8,801	-0.02	3,547	-0.00	-0.02***
	(8,801)	(0.39)	(3,547)	(0.38)	(0.01)
Connectivity Index	8,801	-0.02	3,547	0.04	-0.06***
	(8,801)	(0.40)	(3,547)	(0.41)	(0.01)

Notes: This table shows baseline statistics for treated GPs (whose secretaries received a pay raise) and comparison GPs (whose secretaries did not). Panel A and B show GP statistics. Panel C shows the means of indices created from public goods availability variables in the SHRUG data. We describe how the data is collected in Section A.6. See discussion in Section 2.3.

Table A3: Details on Components of Performance Indicators

Supervisor Assessment	**Self-Reported Assessment**	**Auditor Assessment**	**Citizen Assessment**
Road cleanliness: Road considered clean if supervisor gave assessment of 3 or more (out of 5) and did not find litter, dung, or waste, and made no negative remark. The score is binary. Weight: 13.5%	**Road cleanliness**: On each day, 2 roads should be cleaned. The score is the average of how many roads were cleaned per day on the best 18 days of the month. Weight: 8.2%	**Road cleanliness:** Likert scale (0 to 10) of road cleanliness. Weight: 12.5%	**Cleanliness Satisfaction**: The score is a 5-point Likert scale (1 to 5) which asks about satisfaction regarding road and drain cleanliness. Weight: 16.7%
Drain cleanliness: Drain considered clean if supervisor gave assessment of 3 or more (out of 5) and classified drain as clean, desilted, and maintained, without negative remark. The score is binary. Weight: 13.5%	**Drain cleanliness:** On each day, 2 drains should be cleaned. The score is the average of how many drains were cleaned per day on the best 18 days of the month. Weight: 8.2%	**Drains cleanliness:** Likert scale (0 to 10) of the the drain cleanliness. Weight: 12.5%	
Institution Cleanliness: Supervisors visit several institutions such as schools and healthcare centers. The score is the proportion of clean institutions out of total institutions visited. Weight: 11.5%	**Institution Cleanliness:** An institution (e.g school or healthcare center) should be cleaned daily. Proportion of institutions cleaned among best 18 days. Weight: 7.1%	**Institution Cleanliness:** Likert scale (0 to 10) of the cleanliness of institutions such schools and healthcare centers. Weight: 12.5%	

(continued below)

Table A3: Details on Components of Performance Indicators (continued)

Supervisor Assessment	Self-Reported Assessment	Auditor Assessment	Citizen Assessment
Compost Score: Measures if compost has been taken place. Full score is given if composting procedure is followed. Half score is given if composting has started, but without proper process. Weight: 7.7%	**Household waste segregation:** Based on whether segregated garbage was collected from at least one household. Average of best 18 days in month. Weight: 4.7%	**Waste Segregation Rating:** Likert scale (0 to 10) of whether there was dry-wet waste segregation at household level and how well these wastes have been collected. Weight: 12.5%	
Streetlight functioning: Proportion of working streetlights. Adjusted to zero if below 0.6. Weight: 15.4%	**Streetlight functioning:** Score determined days when streetlights selected roads are fully functioning. Average of best 18 days in month. Weight: 9.3%	**Streetlight functioning:** Likert scale (0 to 10) which assesses how well streetlights are working. Weight: 12.5%	**Streetlight satisfaction**: Likert scale (1 to 5) about satisfaction with functioning of streetlights. Weight: 16.7%
Water tanks: The share of water tanks deemed clean. Weight: 3.8%	**Water tanks:** Proportion of water tanks cleaned in month. Weight: 2.4%		

(continued below)

Table A3: Details on Components of Performance Indicators (continued)

Supervisor Assessment	Self-Reported Assessment	Auditor Assessment	Citizen Assessment
Nursery infrastructure: Proportion of water source, fencing, and cattle traps availability. Weight: 11.5%	**Nursery infrastructure:** Proportion of water source, fencing, and cattle traps availability. Weight: 7.1%	**Nursery Rating:** Likert scale (0 to 10) which assesses whether or not the plant nursery infrastucture is available and working properly. Weight: 12.5%	**Greenery Satisfaction:** Likert scale (1 to 5) about satisfaction with nurseries and plantations. Weight: 16.7%
Plantation fencing: This score aggregates three aspects of plantation fencing: avenue, community, and institutional fencing. Weight: 3.8%	**Plantation fencing:** Score given if each of the 3 plantation types have fencing Weight: 2.4%		
Government plants: Plant survival rate in avenues, community, and institutions (85%). Overall assessment (15%). Weight: 13.5%	**Government plants:** Plant survival rate in avenues, community, and institutions (58%). Watering of plants (42%). Weight: 14.0%	**Government plants:** Likert scale (0 to 10) of government plants including watering and fencing. Weight: 12.5%	
	NREGA: Number of workdays per active beneficiary. First graded as a percentage of best performing GP in Mandal, later graded as a percentage of the 90th percentile performing GP in Mandal. Weight: 18%. In February 2022 this component was removed.		**NREGA Satisfaction** Likert scale (1 to 5) about satisfaction regarding the implementation of NREGA. Weight: 16.7%

(continued below)

Table A3: Details on Components of Performance Indicators (continued)

Supervisor Assessment	**Self-Reported Assessment**	**Auditor Assessment**	**Citizen Assessment**
Meetings: Gram Sabha meeting (33%) and GP meetings (67%) in current or previous month. Weight: 5.8%	**Administration:** Gram Sabha meeting (16.7%) and GP meetings (8.3%) in current or previous month. Proportional of registers that are up to date (16.7%). Proportions of births and deaths registered and certificates issued within timelines (25%). Proportion of daily sanitation records inspections submitted (16.7%). Cheques issued and bank token generated for all due bills (16.7%). Weight: 14.0%		**Administration:** Likert scale (1 to 5) which asks about satisfaction regarding how regular gram sabha meetings are taken place and how timely citizen services like birth registration is carried out. Weight: 16.7%
	Garbage Transport On each day, garbage is supposed to be transported to a shed. The score is the average of whether or not this was followed for the best 18 days of the month. Weight: 4.7%	**Garbage Transport:** Likert Scale (0 to 10) on how well garbage has been transported away from collection areas. Weight: 12.5%	**Garbage Satisfaction** (likert 1 to 5: Very Poor to Excellent) The score is a 5-point Likert scale (1 to 5) which asks about satisfaction regarding household garbage collection. Weight: 16.7%

Notes: This table shows details about the components and how they are aggregated to performance indicators. For the Supervisor Assessment and Self-Reported Assessment, the weights used to aggregate the different components into one score are the same weights as used by the PRRD Department. Refer to the discussion in Section 3.1.

Table A4: Effect of Salary Increase on Supervisors' Subjective Performance Assessment

	(1) Supervisor Subjective Assessment
Pay Increase x Post	0.00706
	(0.0169)
Observations	77,575
Clusters	12,409
Time FE	Yes
Panchayat FE	Yes
Dep. Var. Mean	0

Notes: This table shows the estimated effects of the salary increase on supervisors' subjective assessments of the secretary. The outcome is expressed as SDs from the mean. Standard errors are clustered at the the GP level. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.1.

Table A5: Effect of Salary Increase on Self-Reported Assessments — Verifiable and Non-Verifiable Score Components

	(1) Self-Reported Assessment No-Overlap	(2) Self-Reported Assessment Overlap	(3) Supervisor No Overlap	(4) Supervisor Overlap
Pay Increase x Post	1.413***	-0.334	-0.816	0.165
	(0.276)	(0.252)	(0.547)	(0.205)
Observations	226,922	89,239	53,220	89,239
Clusters	12,770	12,639	11,377	12,639
Time FE	Yes	Yes	Yes	Yes
Panchayat FE	Yes	Yes	Yes	Yes
Dep. Var. Mean	80.7	78.8	71.3	59.8

Notes: This table illustrates the impact of pay increases on the misreporting of both verifiable and non-verifiable score components. In columns (2) and (4), we present effects on verifiable score components — components that are both self-reported by the secretary and evaluated by the supervisor during her inspection (e.g. cleanliness of a particular street). In Column (2), the dependent variable is the self-reported score for these overlapping score components, and in Column (4), it is the supervisor evaluation for these components. Specifically, for each month, these scores are created as the percentage of the maximum possible total, where the total depends on which sub-components between the self-reported score and the supervisor score are both present and non-missing. In columns (1) and (3), we present effects on "non-overlapping" score components — score components where the variable reported by the secretary cannot be directly verified by the supervisor (e.g. frequency of garbage collection or drain cleaning). In Column (1), we report the effects of the salary increase on self-reported scores for non-overlapping components, while in Column (3) we report supervisor evaluations of non-overlapping components (these are things assessed by the supervisor but that the secretary is not required to report on in their daily submission). These non-overlapping scores are also calculated as the percentage of the max possible total. The reason for the absence of certain sub-components in certain months, is due to data availability of the sub-components in that particular month or can be caused by incomplete supervisor assessments during a particular supervisor inspection. Results are discussed in Section 4.2.

Table A6: Traditional Heterogeneity — Effect of Salary Increase on Retention

	Employed						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pay Increase x Post x Heterogeneity		0.001	-0.003	-0.034***	0.011	0.012	0.007
		(0.009)	(0.012)	(0.010)	(0.017)	(0.019)	(0.006)
Pay Increase x Post	0.010**	0.003	-0.002	0.008*	0.011**	0.010**	0.011**
	(0.004)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Post x Heterogeneity		0.005	-0.006	0.003	-0.035**	-0.004	
		(0.008)	(0.010)	(0.009)	(0.016)	(0.018)	
Observations	45,724	44,180	44,048	44,048	30,788	30,788	30,788
Clusters	11,431	11,045	11,012	11,012	7,697	7,697	7,697
Employee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneity		Above Med.	Caste	Female	Dissatisfied	Salary	Promotion

Note: This table shows treatment effect heterogeneity on stay rates. We examine heterogeneity based on five characteristics: "Above Med." corresponds to whether or not a secretary had an above median baseline score in their supervisor assessment score. "Caste" corresponds to whether or not a secretary belongs to a scheduled caste. "Female" corresponds to whether or not a secretary is female. "Dissatisfied" corresponds to whether or not a secretary indicated that they were dissatisfied in the baseline secretary survey. "Salary" corresponds to whether or not the secretary complained about their salary, remuneration, etc during the baseline survey. The coefficient of interest is the triple interaction, i.e. Pay Increase x Post x Heterogeneity. The corresponding DiD figure is 3b. Refer to the discussion in Section 4.3.

Table A7: Robustness of Baseline Results to Time Trends

Panel A: Linear Time Trends	(1) Supervisor Assessment	(2) Auditor Assessment	(3) Citizen Assessment	(4) Effort Metric
Pay Increase x Post	-0.0277	0.5627	-0.1620	-0.0296
	(0.2792)	(0.9529)	(0.1401)	(0.2669)
Time x Pay Increase	-0.0073	-0.0614	0.1033	-0.0376***
	(0.0222)	(0.1232)	(0.0647)	(0.0142)
Observations	105,732	12,221	9,512	459,720
Clusters	12,678	4,851	3,114	12,770
Panchayat FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Dep. Var. Mean	64.7	75.7	.019	66.634
Panel B: Quadratic Time Trends				
	(1) Supervisor Assessment	(2) Auditor Assessment	(3) Citizen Assessment	(4) Effort Metric
Pay Increase x Post	0.2965	1.6734	-0.1614	-0.0296
	(0.4042)	(1.0975)	(0.1402)	(0.2669)
Time^2 x Pay Increase	0.0028	0.0522**	-0.0194	0.0002
	(0.0028)	(0.0239)	(0.0321)	(0.0009)
Time x Pay Increase	-0.0914	-0.8985**	0.3606	-0.0482
	(0.0902)	(0.4217)	(0.4293)	(0.0439)
Observations	105,732	12,221	9,512	459,720
Clusters	12,678	4,851	3,114	12,770
Panchayat FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Dep. Var. Mean	64.7	75.7	.019	66.634

Note: This table shows the robustness of our results to controlling for time trends. We estimate Equation (1) and present DiD estimates. The dependent variables are supervisor performance evaluation (Column 1), auditor performance evaluation (Column 2), citizen evaluation (Column 3) and daily reporting rate, our effort proxy (in Column 4). Panel A controls for linear time trends while Panel B controls for quadratic time trends. Results are discussed in Section 4.4.

Table A8: Effect of Salary Increase on Subjective Performance Evaluation

	(1) Supervisor Overall Assessment	(2) Auditor Overall Assessment	(3) Citizen Overall Assessment
Pay Increase x Post	0.0299*	0.458	-0.0639
	(0.0170)	(0.635)	(0.0709)
Observations	81,526	12,213	9,111
Clusters	12,447	4,847	3,013
Time FE	Yes	Yes	Yes
Panchayat FE	Yes	Yes	Yes
Dep. Var. Mean	0	74.932	0

Notes: This table shows the estimated effects of the salary increase on supervisors', auditors' and citizens' overall evaluation of the village. In both columns 1 and 3, the outcomes are expressed as SDs from the mean. The outcome in Column (2) is expressed as a percentage out of 100. Standard errors are clustered at the the employee level. ***p<0.01; **p<0.05; *p<0.1. Figure A7 shows the timelines for these variables. Figure A8 shows the DiD figures for these variables. See discussion in Section 4.4.

Table A9: Robustness of Baseline Results to Mandal x Month Fixed Effects

	(1) Supervisor Assessment	(2) Auditor Assessment	(3) Citizen Assessment	(4) Effort Metric
Pay Increase x Post	-0.0802	-1.016*	-0.0222	-0.626***
	(0.178)	(0.552)	(0.0987)	(0.157)
Observations	105,285	10,717	9,177	459,648
Clusters	12,666	4,181	3,004	12,768
Time FE	Yes	Yes	Yes	Yes
Panchayat FE	Yes	Yes	Yes	Yes
Mandal-Time FE	Yes	Yes	Yes	Yes
Dep. Var. Mean	64.7	75.61	.029	66.637

Notes: This table presents DiD coefficients from estimating Equation (1). However, unlike the baseline results, which only include GP and month fixed effects, in this regression we additionally include mandal x month fixed effects. In Column (1), the outcome variable is performance as measured by supervisor evaluations, in Column (2), it is auditor evaluation scores, in Column (3), it is citizen evaluation scores, and in Column (4) it is the daily reporting rate, our proxy for effort. Standard errors are clustered at the GP level. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.1.

Table A10: Impact of Salary Increase for GPs with no change in secretary

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Supervisor Assess.	Auditor Assess.	Citizen Assess.	Effort Metric	Self-reported Assess.	Score Inflation (%)	Sarpanch Self-Dealing (%)	Secretary Self-Reimb. (%)
Pay Increase x Post	0.174	0.0392	-0.0451	-0.678***	0.348*	-0.473	0.151	0.0195
	(0.332)	(0.823)	(0.105)	(0.237)	(0.196)	(0.461)	(0.377)	(0.0183)
Observations	62,134	7,095	5,494	269,172	142,063	52,432	83,549	145,134
Clusters	7,425	2,793	1,795	7,477	7,477	7,406	6,213	7,477
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panchayat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep. Var. Mean	64.8	75.78	.037	67.2	76.8	19.1	8.6	.04

Notes: This table shows the estimated effects of the salary increase on the performance of the secretaries who stayed in the same GP from May 2020 to Mar 2023. We present DiD coefficients from estimating Equation (1). In Column (1) the outcome variable is the performance score constructed from the supervisors' assessments, in Column (2), it is the performance score constructed from assessments by state-level auditors. For Column (3), it is the score constructed from the citizen assessments, and in Column (4) it is the percentage of days the secretary files the required daily sanitation report, which we consider a proxy for effort. In Column (5), the outcome variable is the performance index constructed from the self-reported performance indicators, expressed as a percentage out of 100. In Column (6), the outcome variable is score inflation, as defined in Section 3.3.1, expressed as a percentage out of 100. In Column (7) shows the changes over time for the share of multipurpose workers hired sharing a name with the GP Sarpanch, Column (8) shows the share of GP expenditure that was spent in reimbursing the GP secretary, expressed as the percent of total monthly expenditure. Standard errors are clustered at the GP level. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 4.4.

Table A11: Effects of Relative Pay on Secretary Performance — Supervisor Score, Auditor Score, Citizen Score, and Effort (Grade IV vs Other Grades)

	(1) Supervisor Assessment	(2) Auditor Assessment	(3) Citizen Assessment	(4) Effort Metric
Grade 4 PS x Post	0.0440	-2.181**	-0.0718	0.169
	(0.435)	(1.060)	(0.135)	(0.343)
Observations	31,129	3,730	2,831	132,768
Clusters	3,662	1,441	930	3,688
Time FE	Yes	Yes	Yes	Yes
Panchayat FE	Yes	Yes	Yes	Yes
Dep. Var. Mean	64.5	75.7	.013	66.39

Notes: This table shows the estimated effects of the salary increase on the performance of the Grade IV secretaries who used to be in the second lowest paid position but who after the pay increase are in the shared lowest position in terms of pay. The table is displaying the DiD coefficient β from estimating Equation (1). In Column (1) the outcome variable is the performance score constructed from the supervisor evaluations, in Column (2), it is the performance score constructed from auditor evaluations. For Column (3), it is the score constructed from citizen assessments, and in Column (4) it is the daily reporting rate, i.e. how regularly secretaries file the required daily sanitation report, which is a proxy for effort. Standard errors are clustered at the GP level. ***p<0.01; **p<0.05; *p<0.1. See discussion in Section 5.3.

Table A12: Predictions by Forecasters' Subject Expertise, Contextual Knowledge, and Education

Panel A. Respondents Familiar with Literature (n = 29)			
Outcome	Median	Mean	Std. Dev.
Supervisor Assessment	0.25	0.27	(0.22)
Self-reported Assessment	0.31	0.33	(0.22)
Secretary Corruption Proxy	4.11	3.90	(3.24)
	Share of Predictions (%)		
	Worsen	Unchanged	Improves
Selection	10.30	13.80	75.90
	Decreases	Unchanged	Increases
Secretary Corruption Proxy	72.40	17.20	10.30
Sarpanch Corruption Proxy	48.30	27.60	24.10
Panel B. Respondents Familiar with India (n = 11)			
Outcome	Median	Mean	Std. Dev.
Supervisor Assessment	0.22	0.18	(0.26)
Self-reported Assessment	0.30	0.30	(0.30)
Stay-rate	5.70	4.62	(3.75)
	Share of Predictions (%)		
	Worsen	Unchanged	Improves
Selection	18.20	18.20	63.60
	Decreases	Unchanged	Increases
Secretary Corruption Proxy	81.80	9.10	9.10
Sarpanch Corruption Proxy	45.50	36.40	18.20
Panel C. Respondents with a PhD (n = 33)			
Outcome	Median	Mean	Std. Dev.
Supervisor Assessment	0.22	0.24	(0.16)
Self-reported Assessment	0.30	0.31	(0.19)
Stay-rate	3.17	3.68	(3.09)
	Share of Predictions (%)		
	Worsen	Unchanged	Improves
Selection	21.20	15.20	63.60
	Decreases	Unchanged	Increases
Secretary Corruption Proxy	81.80	9.10	9.10
Sarpanch Corruption Proxy	51.50	30.30	18.20

Notes: This table presents forecasts from our Social Science Prediction Platform survey. Panel A shows the predictions of respondents who say they are extremely familiar or very familiar with the existing literature on public sector pay. Panel B shows the predictions of respondents stating that they are extremely familiar or very familiar with the Indian context. Panel C shows the predictions of respondents who have a PhD .Predictions for supervisor assessment, and self-reported assessment are in SDs from the mean. Predictions for stay rate are in percentage points. For the average predictions of all forecaster, refer to Table 4. Results are discussed in Section 4.5.

Table A13: Traditional Heterogeneity — Effect of Salary Increase on Performance (Supervisor and Auditor Evaluations)

Panel A:	Supervisor Assessment						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pay Increase x Post x Heterogeneity		0.449	1.315**	-0.069	-0.445	-0.255	-0.268
		(0.405)	(0.611)	(0.542)	(0.848)	(0.982)	(0.397)
Pay Increase x Post	-0.144	-0.271	-0.357	-0.096	-0.391	-0.401	-0.378
	(0.231)	(0.315)	(0.264)	(0.279)	(0.296)	(0.294)	(0.283)
Post x Heterogeneity		-11.205***	-0.954*	-0.062	1.198	0.323	
		(0.341)	(0.515)	(0.473)	(0.755)	(0.923)	
Observations	111,223	110,620	105,826	105,826	84,353	84,353	84,353
Clusters	12,697	12,573	12,107	12,107	9,425	9,425	9,425
Panchayat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneity		Above Med.	Caste	Female	Dissatisfied	Salary	Promotion
Panel B:	**Auditor Assessment**						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pay Increase x Post x Heterogeneity		1.191	1.556	-1.163	1.939	3.504	-1.613*
		(1.130)	(1.586)	(1.422)	(2.069)	(2.645)	(0.921)
Pay Increase x Post	0.171	-0.440	-0.024	0.610	0.061	-0.033	0.415
	(0.566)	(0.797)	(0.644)	(0.667)	(0.767)	(0.754)	(0.730)
Post x Heterogeneity		-1.441	-1.286	0.036	-3.302*	-4.137	
		(0.969)	(1.393)	(1.273)	(1.883)	(2.521)	
Observations	12,221	12,119	11,580	11,580	9,257	9,257	9,257
Clusters	4,851	4,810	4,613	4,613	3,638	3,638	3,638
Panchayat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneity		Above Med.	Caste	Female	Dissatisfied	Salary	Promotion

Note: This table shows treatment effect heterogeneity on performance as measured by supervisor and auditor evaluations. Panel A shows the heterogeneous effects of the pay increase on the Supervisor Assessment Score. Panel B shows treatment effect heterogeneity on Auditor Assessment. We examine heterogeneity based on five characteristics: "Above Med." corresponds to whether or not a secretary had an above median baseline score in their supervisor assessment score. "Caste" corresponds to whether or not a secretary belongs to a scheduled caste. "Female" corresponds to whether or not a secretary is female. "Dissatisfied" corresponds to whether or not a secretary indicated that they were dissatisfied in the baseline secretary survey. "Salary" corresponds to whether or not the secretary complained about their salary, remuneration, etc during the baseline survey. "Promotion" corresponds to whether or not the secretary complained about promotion, regularization, etc during their baseline survey. For the characteristic "Promotion", the interaction term *Post* × *Commented Promotion* is excluded due to collinearity. The coefficient of interest is the triple interaction, i.e. Pay Increase x Post x Heterogeneity. The corresponding DiD figure is A11. Refer to the discussion in Section 5.

Table A14: Traditional Heterogeneity — Effect of Salary Increase on Performance (Citizen Evaluations and Effort)

Panel A:			Citizen Assessment				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pay Increase x Post							
x Heterogeneity		0.003	-0.162	0.083	0.185	0.126	0.006
		(0.146)	(0.181)	(0.177)	(0.242)	(0.230)	(0.108)
Pay Increase x Post	0.022	0.030	0.035	-0.017	0.020	0.027	0.037
	(0.073)	(0.103)	(0.085)	(0.088)	(0.094)	(0.094)	(0.089)
Post x Heterogeneity		-0.000	0.325**	-0.040	-0.238	-0.123	
		(0.126)	(0.154)	(0.158)	(0.219)	(0.206)	
Observations	9,512	9,445	9,047	9,047	7,099	7,099	7,099
Clusters	3,114	3,085	2,957	2,957	2,312	2,312	2,312
Panchayat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneity		Above Med.	Caste	Female	Dissatisfied	Salary	Promotion
Panel B:			Effort Metric				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pay Increase x Post							
x Heterogeneity		0.199	-0.171	0.272	1.346**	2.178***	-0.453*
		(0.349)	(0.484)	(0.409)	(0.594)	(0.596)	(0.242)
Pay Increase x Post	-0.707***	-0.805***	-0.734***	-0.772***	-0.452**	-0.462**	-0.224
	(0.174)	(0.255)	(0.197)	(0.210)	(0.192)	(0.194)	(0.186)
Post x Heterogeneity		-0.049	0.072	-1.029***	-1.288**	-2.299***	
		(0.302)	(0.425)	(0.363)	(0.532)	(0.553)	
Observations	459,720	454,356	438,408	438,408	341,064	341,064	341,064
Clusters	12,770	12,621	12,178	12,178	9,474	9,474	9,474
Panchayat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneity		Above Med.	Caste	Female	Dissatisfied	Salary	Promotion

Note: This table shows treatment effect heterogeneity on performance as measured by citizen evaluations and effort. Panel A shows the heterogeneity in the effect of pay increase on the Citizen Assessment. Panel B shows heterogeneity in the effect of higher salaries on the daily reporting rate, our proxy for effort. We examine heterogeneity based on five characteristics: "Above Med." corresponds to whether or not an secretary had a above median baseline score in their supervisor assessment score. "Caste" corresponds to whether or not a secretary belongs to a scheduled caste. "Female" corresponds to whether or not a secretary is female. "Dissatisfied" corresponds to whether or not a secretary indicated that they were dissatisfied in the baseline secretary survey. "Salary" corresponds to whether or not the secretary complained about their salary, remuneration, etc during the baseline survey. "Promotion" corresponds to whether or not the secretary complained about promotion, regularization, etc during their baseline survey. For the characteristic "Promotion", the interaction term *Post* × *Commented Promotion* is excluded due to collinearity. The coefficient of interest is the triple interaction, i.e. Pay Increase x Post x Heterogeneity. The corresponding DiD figure is A11. Refer to the discussion in Section 5.

Table A15: Traditional Heterogeneity — Effect of Salary Increase on Honesty

Panel A:	Self-Reported Assessment						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pay Increase x Post x Heterogeneity		-0.271	-0.074	-0.490	0.270	0.091	0.324
		(0.283)	(0.397)	(0.324)	(0.509)	(0.504)	(0.230)
Pay Increase x Post	0.417***	0.568***	0.441***	0.587***	0.567***	0.598***	0.564***
	(0.141)	(0.210)	(0.160)	(0.172)	(0.171)	(0.173)	(0.165)
Post x Heterogeneity		-0.308	-0.066	-0.090	0.013	-0.068	
		(0.240)	(0.341)	(0.280)	(0.454)	(0.462)	
Observations	242,630	239,799	231,382	231,382	180,006	180,006	180,006
Clusters	12,770	12,621	12,178	12,178	9,474	9,474	9,474
Panchayat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneity		Above Med.	Caste	Female	Dissatisfied	Salary	Promotion
Panel B:	**Score Inflation (%)**						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pay Increase x Post x Heterogeneity		-0.112	-0.897	0.850	1.040	-0.654	-0.326
		(0.635)	(0.848)	(0.767)	(1.144)	(1.187)	(0.527)
Pay Increase x Post	-0.499	-0.451	-0.303	-0.671*	-0.440	-0.223	-0.269
	(0.317)	(0.469)	(0.365)	(0.381)	(0.397)	(0.398)	(0.379)
Post x Heterogeneity		1.398***	1.431**	-0.839	-0.838	0.331	
		(0.540)	(0.726)	(0.677)	(1.023)	(1.101)	
Observations	89,239	88,836	84,896	84,896	67,656	67,656	67,656
Clusters	12,639	12,527	12,051	12,051	9,386	9,386	9,386
Panchayat FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneity		Above Med.	Caste	Female	Dissatisfied	Salary	Promotion

Note: This table shows treatment effect heterogeneity on honesty. Panel A shows heterogeneity in the effect of pay increase on the Self-Reported score. Panel B shows the heterogeneity of the effect of pay increase on inflation. We examine heterogeneity based on five characteristics: "Above Med." corresponds to whether or not a secretary had an above median baseline score in their supervisor assessment score. "Caste" corresponds to whether or not a secretary belongs to a scheduled caste. "Female" corresponds to whether or not a secretary is female. "Dissatisfied" corresponds to whether or not a secretary indicated that they were dissatisfied in the baseline secretary survey. "Salary" corresponds to whether or not the secretary complained about their salary, remuneration, etc during the baseline survey. "Promotion" corresponds to whether or not the secretary complained about promotion, regularization, etc during their baseline survey. For the characteristic "Promotion", the interaction term *Post* × *Commented Promotion* is excluded due to collinearity. The coefficient of interest is the triple interaction, i.e. Pay Increase x Post x Heterogeneity. The corresponding DiD figure is A12. Refer to the discussion in Section 5.

Table A16: Traditional Heterogeneity — Effect on Salary Increase on Corruption

Panel A:	Sarpanch Self Dealing (%)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pay Increase x Post							
x Heterogeneity		0.438	0.461	-0.272	0.869	-1.380	0.792*
		(0.508)	(0.598)	(0.591)	(0.916)	(0.926)	(0.449)
Pay Increase x Post	0.168	-0.023	0.145	0.312	-0.022	0.260	-0.006
	(0.252)	(0.364)	(0.292)	(0.300)	(0.302)	(0.309)	(0.297)
Post x Heterogeneity		-0.424	-0.608	-0.071	-0.473	1.137	
		(0.405)	(0.450)	(0.486)	(0.793)	(0.832)	
Observations	139,045	137,407	133,343	133,343	105,459	105,459	105,459
Clusters	10,503	10,383	10,050	10,050	7,861	7,861	7,861
Employee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneity		Above Med.	Caste	Female	Dissatisfied	Salary	Promotion
Panel B:	Secretary Self Reimbursements (%)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pay Increase x Post							
x Heterogeneity		-0.013	-0.039	0.025	0.046	0.024	-0.020
		(0.024)	(0.032)	(0.023)	(0.053)	(0.046)	(0.013)
Pay Increase x Post	0.010	0.019	0.019	0.004	0.011	0.017	0.020
	(0.012)	(0.018)	(0.012)	(0.014)	(0.013)	(0.015)	(0.014)
Post x Heterogeneity		0.011	0.048	-0.004	-0.025	-0.039	
		(0.019)	(0.029)	(0.016)	(0.044)	(0.044)	
Observations	247,766	244,886	236,313	236,313	184,011	184,011	184,011
Clusters	12,770	12,621	12,178	12,178	9,474	9,474	9,474
Employee FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Heterogeneity		Above Med.	Caste	Female	Dissatisfied	Salary	Promotion

Note: This table shows treatment effect heterogeneity on corruption. Panel A shows heterogeneity in the effect of pay increase on nepotism in hiring, as measured by the share of multipurpose workers who share a name with the Sarpanch. Panel B shows the heterogeneity in the effect of pay increase on Secretary Self-Reimbursements. We examine heterogeneity based on five characteristics: "Above Med." corresponds to whether or not a secretary had an above median baseline score in their supervisor assessment score. "Caste" corresponds to whether or not a secretary belongs to a scheduled caste. "Female" corresponds to whether or not a secretary is female. "Dissatisfied" corresponds to whether or not a secretary indicated that they were dissatisfied in the baseline secretary survey. "Salary" corresponds to whether or not the secretary complained about their salary, remuneration, etc during the baseline survey. "Promotion" corresponds to whether or not the secretary complained about promotion, regularization, etc during their baseline survey. For the characteristic "Promotion", the interaction term *Post* × *Commented Promotion* is excluded due to collinearity. The coefficient of interest is the triple interaction, i.e. Pay Increase x Post x Heterogeneity. The corresponding DiD figure is A12. Refer to the discussion in Section 5.

Table A17: Generic ML Heterogeneity — Correlation Among Predicted Treatment Effects

	(1) Supervisor TE	(2) Effort TE	(3) Corruption TE	(4) Retention TE
Effort TE	0.0695**		-0.0452	0.000389
	(0.0279)		(0.0332)	(0.000312)
Retention TE	1.991*	0.816	-1.873	
	(1.160)	(0.655)	(1.499)	
Corruption TE	-0.0212**	-0.00562		-0.000111
	(0.00938)	(0.00410)		(0.0000888)
Supervisor TE		0.0156**	-0.0383**	0.000213*
		(0.00623)	(0.0169)	(0.000123)
Observations	7,163	7,163	7,163	7,163

Note: This table shows the correlation between the predicted treatment effects for different outcomes. *Supervisor TE* denotes each secretary's median predicted treatment effect on supervisor evaluations using Chernozhukov et al. (2018)'s GenericML approach. Specifically, from each sample split, we obtain a predicted TE for each secretary. The median predicted TE is simply the median of the predicted TEs across all sample splits. . *Effort TE and Retention TE* are variables that respectively denote each secretary's median predicted TE on the reporting rate and stay rate. *Corruption TE* denotes each secretary's median predicted TE on self-dealing (share of GP expenditure paid to the secretary for personal expenses). Refer to the discussion in Section 5.

Table A18: Predicted Treatment Effects on Combination of Outcomes

Category	**Performance**	**Stay rate**	**Effort**	**Honesty**	**#(+)**	**#(-)**	**Share of secretaries**	
							Overall	Salary complaint
Clearly Positive	+	+	+	+	4	0	0%	0%
Weakly Positive	+ or 0	+ or 0	+ or 0	+ or 0	> 0	0	21.4%	29.1%
Tradeoff	+, 0 or -	+, 0 or -	+, 0 or -	+, 0 or -	> 0	> 0	3.7%	4.3%
Weakly Negative	- or 0	- or 0	- or 0	- or 0	0	> 0	15.6%	11.3%
Clearly Negative	-	-	-	-	0	4	0%	0%
No effect	0	0	0	0	0	0	59.3%	55.3%

Note: This table summarizes the predicted treatment effects across different outcomes. The first row (*Very Good*) refers to secretaries for whom predicted treatment effects on all four outcomes — performance, stay rates, effort and honesty — are positive and statistically significant. The second row (*Weakly Good*) refers to secretaries for whom higher salaries are predicted to have positive and significant effects on at least one outcome and have negative and significant on none. The third row (*Tradeoffs*) refers to secretaries where predicted treatment effects are positive and significant for at least one outcome and are negative and significant for at least one outcome. *Weakly Negative* refers to secretaries with at least one negative and significant predicted treatment effect and no positive and significant treatment effects. *Clearly Negative* refers to secretaries predicted to have negative and significant treatment effects on all four outcomes. *No effect* refers to secretaries predicted to have null effects on all four outcomes. The final three columns indicate the share of secretaries in each category. The *Overall* column contains shares on the full sample of secretaries. The *Salary Complaint* complaint shows shares for secretaries who had a baseline salary complaint, while *Top performer* shows shares for secretaries in the top quartile of baseline supervisor evaluation scores. Refer to the discussion in Section 5.

E Social Science Prediction Platform Survey

We collect expert forecasts via the Social Sciences Prediction Platform (SSPP), whose mission is to "enable the systematic collection and assessment of expert forecasts of the effects of untested social programs."[26] Our main goal with the SSPP survey was to understand how our findings compared with the priors of scholars who are informed about the Indian context and academic literature on public sector pay. Participants were provided with background information about the policy reform and the geographical context. After being asked a round of demographic and background questions the respondents were asked to predict the effect of the salary increase on:

- Secretaries' performance as assessed by supervisor evaluations on a sliding scale from -1 to 1 standard deviations
- Self-reported performance on a sliding scale from -1 to 1 standard deviations
- Stay rate in government service on a scale from -10 to +10 percentage points
- The quality of secretaries who remain in local government service. The options provided as answers to this questions were directional: improve, worsen, or no effect
- Corruption related to hiring and the awarding of contracts. The options provided as answers to this questions were directional: increase, decrease, or no effect
- Corruption as measured by self-dealing by the GP secretary. The options provided as answers to this questions were directional: increase, decrease, or no effect

For each question, the respondents were also asked to provide their confidence in the prediction on a five-point Likert scale from "Extremely Confident" to "Not at All Confident".

[26] For more information see: https://socialscienceprediction.org/