Automated Assignment and Bureaucrat Productivity: Evidence from a Field Experiment in Peru

Alipio Ferreira   Sandra Cermeño León

June 11, 2026

Abstract

Governments increasingly use algorithmic tools to manage bureaucratic production, replacing or supporting human discretion. We study this distinction in a field experiment with Peru's environmental enforcement agency. Analysts were randomly assigned to receive enforcement case lists generated either by managers or by an algorithm. Independently, suggested completion dates were randomized at the case level. Replacing managerial assignment with reduced timely case completion by 5 percentage points, while suggested completion dates increased timely completion by approximately 4 percentage points. Managers use discretion to match complex cases to experienced analysts, and the automated assignment reduces productivity by breaking this link.

JEL Classification Codes: D73, H83, Q58, M11, C93

Keywords: algorithms; bureaucratic productivity; managerial discretion; environmental enforcement

[1]Alipio Ferreira: Southern Methodist University, alipioferreira@smu.edu; Sandra Cermeño León: Innovations for Poverty Action, scermeno@poverty-action.org. We thank Anne Brockmeyer, Nathaniel Pattison, Timothy Salmon, and participants at the RIDGE LACEA BRAIN conference for helpful comments. We gratefully acknowledge the support at IPA Peru from Guillermo Delgado and Fernando Tavara, at J-PAL LAC from César Huaroto, and at OEFA from Elsa López, Gustavo Cuellar, Miriam Rocío Roncal, and Ricardo Machuca. This study was pre-registered in AsPredicted under trial #156452 and in the AEA Registry with RCT ID AEARCTR-0018756. All errors are our own.

1 Introduction

In many organizations, managers must allocate tasks across heterogeneous workers. Assignment decisions affect organizational output because tasks differ in importance and complexity, and workers differ in skill, speed, and comparative advantage. As organizations scale, this assignment problem can quickly become too complex to solve without formal tools. A growing body of evidence suggests that algorithmic interventions can generate positive social returns in prediction and selection settings (Ludwig, Mullainathan, and Rambachan 2024). Yet in some settings, managerial discretion enables the flexible use of soft information about workers and tasks that cannot easily be encoded in administrative data. The level of managerial discretion is particularly important in government bureaucracies because of corruption concerns and the lack of strong productivity incentives (Finan, Olken, and Pande 2017). Whether automated assignment rules outperform managerial discretion is therefore a context-dependent empirical question.

This paper provides experimental evidence on the effects of algorithmic task assignment in a large public-sector bureaucracy. In partnership with Peru's national environmental enforcement agency (OEFA), we study the introduction of an automated assignment system in the unit responsible for processing sanctioning cases against regulated firms. These cases are legally complex, subject to strict statutory deadlines, and vary widely in urgency and difficulty. Prior to the intervention, managers allocated cases discretionarily across legal analysts. Given the volume and diversity of cases, purely discretionary assignment risked producing unbalanced workloads and failing to systematically account for cases nearing their statutory prescription deadlines. In 2023, OEFA introduced a system that automated assignment to analysts, replacing managerial discretion in the selection process (hard automation). Moreover, the system also added suggested completion dates for task completion as a behavioral nudge aimed at structuring analysts' workflow (soft automation).

This environment reflects conditions common in many organizations where algorithmic management tools are being introduced: workers are heterogeneous in speed and quality; tasks are procedurally constrained and legally time-bound; backlogs create acute pressure to prioritize; and key dimensions of worker skill and fit are difficult to observe or encode in administrative data. These features are not unique to public enforcement. They also characterize private-sector professional services such as legal services, consulting, auditing, and compliance, where case complexity and risk vary widely and supervisors often rely on soft information about who can handle which work under tight deadlines.

We exploit a pre-registered randomized experiment with two cross-randomized treatments. At the analyst level, analysts were assigned either to continue receiving cases from their managers or to receive the algorithm's case lists. The automated assignment of cases was defined through a linear optimization algorithm that selects and orders cases given the capacity constraints of analysts and the expected completion times of each procedural stage; a formulation known as the knapsack problem, which has received limited empirical attention in the management literature (Schiffels, Fliedner, and Kolisch 2013, Pape, Kavadias, and Sommer 2024). At the case level, some cases displayed suggested completion dates. This design lets us estimate the effect of replacing managerial discretion in assignment separately from the effect of deadline-based scheduling cues. Using detailed administrative data on 1,405 enforcement cases processed by 95 analysts across eight economic sectors, we track whether cases were started, completed on time, expired, and ultimately resulted in penalties.

We document three main results. First, automated assignment reduced timely case completion: analysts whose cases were selected by the algorithm were approximately 5 percentage points less likely to conclude a case on time relative to those whose cases were assigned by managers. This decline is large relative to a control-group completion rate of about 40 percent, and becomes more pronounced as the backlog accumulates over time. Second, suggested completion dates increased timely completion by about 4 percentage points, indicating that deadline information influences behavior even though it does not offset the negative effect of algorithmic case selection. This effect is concentrated among cases assigned by managers, suggesting that scheduling cues are more effective when managers retain discretion in case selection. Third, neither intervention affected enforcement quality: fines, appeal rates, and appeal outcomes were statistically indistinguishable across groups.

We find no evidence that either intervention affected the pecuniary fines assigned to cases, suggesting that the observed effects on completion do not reflect a trade-off between speed and quality. Moreover, using additional administrative data on case appeals, we find no differences in the probability and outcomes of appeals, suggesting that the quality of the legal analysis was not affected by the interventions.

These results highlight a fundamental trade-off in algorithmic management. The optimization algorithm produces more transparent and equalized workloads, which may improve worker satisfaction and fairness perceptions. But the algorithm also replaces managerial discretion that may incorporate valuable soft information about worker capacity, sector familiarity, and writing quality—dimensions that are hard to encode in administrative data. When worker heterogeneity is high and private information matters, fairness or equalization constraints can be productivity-reducing. In contrast, light-touch process structuring, such as suggested milestone dates, can improve output while preserving discretion, pointing to the value of hybrid designs rather than fully automated assignment.

The results should be interpreted with attention to context. The experiment identifies the effect of replacing this specific discretionary system with this specific algorithmic assignment regime in a setting characterized by high workload, temporary staffing expansion, and managers who likely held valuable soft information about worker-case fit. It does not speak to the effects of algorithmic support more broadly; designs that combine algorithmic recommendations with managerial override authority may yield different outcomes.

The assignment tool combined agency-set capacity parameters with case-level information on complexity, prescription risk, and expected processing times for the main procedural milestones. In the implementation files used for the experiment, OEFA generated one hundred candidate seeded allocations and selected the one with the highest projected productivity for use in the field. Relative to discretionary assignment, the treatment therefore changes two features of production at once: it removes managerial judgment from worker-case matching and introduces a codified assignment-and-scheduling rule that compresses the distribution of workload across analysts. Suggested completion dates constitute an additional, separately randomized intervention layered on top of this new assignment rule.

This paper contributes to several strands of the literature. At the broadest level, it speaks to the economics of organizations: specifically, the question of how to match workers to tasks and the value of managerial autonomy (Kellogg, Valentine, and Christin 2020, Benlian et al. 2022). Algorithmic and rule-based assignment can improve perceived fairness and productivity in some settings: for example, Bai et al. 2022 find that workers perceived algorithmic assignment as fairer than human assignment and became more productive, while discretion can affect productivity and worker welfare (Bartling, Fehr, and Schmidt 2013).

At the same time, replacing human judgment may reduce performance when discretion embeds valuable information. In environmental enforcement, Kang and Silveira (2021) show that regulator discretion can improve enforcement by allowing punishments to be tailored to regulated entities' attributes. In contrast, reducing discretion may be beneficial when it curbs favoritism, leakage, or weak incentives, as shown in settings such as cash-transfer targeting, e-governance, anti-corruption monitoring, and bureaucratic incentives (Haseeb and Vyborny 2022, Banerjee et al. 2020, Olken 2006, Olken 2007, Khan, Khwaja, and Olken

2016, Khan, Khwaja, and Olken 2019). The optimization problem at the core of our intervention (selecting and ordering cases under capacity and deadline constraints) is a variant of the knapsack problem broadly applicable to professional service organizations but that has received limited empirical attention (Schiffels, Fliedner, and Kolisch 2013, Pape, Kavadias, and Sommer 2024).

The paper also contributes to the literature on algorithmic management in bureaucracies (Finan, Olken, and Pande 2017). Several recent papers study settings where an algorithm ranks or selects cases for human review, such as bail decisions (Kleinberg et al. 2018) and tax audit selection (Bachas et al. 2025). A conceptually important difference in our problem is that it embeds a matching problem between cases and workers, introducing an additional layer of complexity that requires information about worker capacity, skill, and comparative advantage. The closest studies to ours examine different forms of algorithmic management, rule-based assignment, or discretionary case selection. Margalit and Raviv (2024) study how workers respond to human versus algorithmic management in an online labor-market field experiment. Bachas et al. (2025) compare tax audits selected by inspectors with audits selected by a transparent risk-scoring algorithm in Senegal. Duflo et al. (2018) compare random inspections with discretionary regulatory targeting in Indian environmental enforcement. Across these settings, removing or weakening discretionary human judgment can reduce performance, especially when human decision-makers possess relevant information that is not fully captured in the data. Our paper extends this literature by testing two types of automation simultaneously (a hard intervention that removes managerial discretion and a soft behavioral nudge) in a legally constrained enforcement setting implemented at full scale in a major public agency.

A growing literature has documented benefits from structuring bureaucratic workflow through suggested, non-binding completion dates. Campbell and Gingrich 1986 show that, for complex tasks, participation in discussing and setting completion targets improved performance relative to assigned targets of similar difficulty. Deadlines can shape attention and induce earlier action (Altmann, Traxler, and Weinschenk 2022), though they may induce procrastination when they are looser than workers expected (Knowles et al. 2022).

The remainder of this paper is organized as follows. Section 2 describes the Peruvian environmental enforcement structure and the organization where the study was conducted. Section 3 discusses the intervention's theory of change. Section 4 details the intervention and the experimental design, and Section 5 discusses the empirical analysis. Section 6 describes the data and descriptive statistics. Section 7 presents the results, Section 8 presents evidence on mechanisms and a discussion. Section 9 concludes.

2 Background

The Agency for Environmental Evaluation and Enforcement (OEFA, for its acronym in Spanish) is Peru's national authority responsible for supervising and enforcing environmental regulations. The agency is organized into directorates that specialize in environmental regulation, monitoring, inspections, and sanctioning. At the sanctioning stage, the Directorate of Enforcement and Incentive Application (DFAI, for its acronym in Spanish) determines sanctions and applies penalties, including fines. DFAI acts upon reports issued by the inspection directorates, which identify presumed infractions and recommend corrective actions. Once a case is referred to DFAI, enforcement becomes a desk-based legal procedure in which analysts establish the relevant infractions, propose fines, and draft other sanctions. Regulated entities may then accept the sanction or appeal. As such, DFAI constitutes the core of OEFA's enforcement authority.

The sanctioning process at OEFA, Administrative Sanctioning Procedure (PAS, for its acronym in Spanish), is composed of three steps. Upon receipt of a supervision report, the first formal step is the issuance of a Sub-Directorial Resolution (RSD), which brings charges against the regulated entity. After the regulated entity has an opportunity to respond, the instructing authority prepares a Final Instruction Report (IFI), which analyzes the case, assesses responsibility, and recommends whether the decision authority should impose a sanction and, where applicable, a fine. The decision authority then issues a Directorial Resolution (RD), determining administrative liability and the amount of the fine or ordering the case archived. The sample case files also show that, when the IFI recommends responsibility, the regulated entity may submit an additional round of descargos to the decision authority before the RD is issued. The whole process has a maximum duration of 9 months from the notification of the charges (RSD), extendable once for up to 3 additional months by a duly justified resolution issued before the original deadline, otherwise the procedure expires ("caducidad") and must be archived.

Cases are assigned to legal analysts specialized by economic sector. Traditionally, team leaders, or "managers," allocated cases discretionarily across analysts, often producing uneven workloads. Moreover, the volume of cases is high, with many cases being at risk of prescription and in need of urgent analysis. The inability to effectively manage the workload on time may weaken OEFA's ability to deter non-compliance and compromise the credibility of environmental enforcement.

To address these concerns, OEFA proposed experimenting with an algorithm that would assign cases to analysts without managerial input, with the goal of increasing productivity, reducing the risk of case expiration, and guaranteeing a more balanced workload across workers. In the first half of 2023, DFAI developed an assignment algorithm with the goal of automating the assignment of thousands of cases to be conducted in the second half of the year. The algorithm assigned cases using information on case complexity, prescription dates, and expected days required for the RSD, IFI, and RD stages. It also provided suggested dates of completion for the cases based on their ordering and the capacity constraint of analysts.

Importantly, DFAI was seeking to reduce what it saw as a large case backlog during the second half of 2023. To boost this effort, DFAI hired dozens of temporary analysts to support case analysis, thus expanding the unit's staff for approximately half a year. This hiring surge increased heterogeneity in experience, sector familiarity, and speed of drafting procedural documents. At the same time, many relevant dimensions, such as writing quality, knowledge of sector-specific jurisprudence, and reliability under time pressure, were not directly measurable in the administrative data available to the algorithm but were partially known by team leaders through day-to-day interaction.

3 Conceptual Framework

The theory of change behind algorithmic assignment rests on two potential channels through which it may improve productivity. First, by equalizing workloads across analysts, the algorithm may reduce the perceived unfairness associated with unbalanced case distributions—a mechanism documented in settings where algorithmic assignment improved fairness perceptions and productivity (Bai et al. 2022, Pechová, Volfová, and Jírová 2023, Beer, Qi, and Rios 2025). Workload equalization is therefore a potential pathway through which the algorithm may generate output gains, independent of any effect on case prioritization. Second, by encoding prescription-risk prioritization, the algorithm systematically directs analysts toward cases at greatest risk of statutory expiration, which may reduce case loss and increase timely completion.

The potential cost of algorithmic assignment is the loss of managerial soft information. A hypothesis for why discretion should be retained is that managers who interact closely with their analysts possess private knowledge about writing quality, sector familiarity, reliability under time pressure. That knowledge enables them to direct complex or urgent cases toward the analysts best equipped to handle them. This is consistent with evidence from tax enforcement and environmental inspections (Bachas et al. 2025, Duflo et al. 2018). At the same time, managerial discretion may itself be subject to constraints: in a high-volume setting, cognitive and time limitations may prevent optimal matching, and discretionary decisions may also reflect favoritism or other non-productivity considerations. Whether the informational advantage of managers outweighs the productivity gains from equalization and prescription-risk prioritization is the central empirical question the experiment is designed to answer.

These considerations imply that algorithmic and discretionary assignment represent fundamentally different information environments. Discretionary assignment exploits tacit managerial knowledge about who is best suited to which case; algorithmic assignment replaces that judgment with formalized optimization rules that explicitly encode fairness and expiration-risk constraints.

On top of the question about the value of discretionary assignment (hard automation), the experiment also tests the effect of suggested completion dates (soft automation). These dates constitute a behavioral nudge: they do not constrain analyst choices but provide scheduling information intended to help analysts organize their work and prioritize cases. Such nudges may also induce procrastination if the suggested dates are looser than workers expected. The effect of scheduling cues may differ across assignment regimes: suggested dates may be more effective under discretionary assignment, where they can complement managerial judgment about case-worker fit, than under algorithmic assignment.

4 Intervention and Design

This paper evaluates two distinct automation treatments, implemented as a cross-randomized field experiment within OEFA's sanctioning unit. The first, which we call hard automation, replaces managerial discretion in case assignment: instead of team leaders allocating cases to analysts, an optimization algorithm generates the assignment list. The second, which we call soft automation, preserves the assignment decision but structures the information available to analysts: the algorithm produces suggested milestone dates for completing each procedural stage, displayed for a randomly selected subset of cases. These two treatments differ fundamentally in the degree to which they displace human judgment. Hard automation eliminates the team leader's role in matching cases to workers. Soft automation leaves that matching unchanged while providing analysts with scheduling information intended to help them organize their work.

The hard automation system generated case allocations using expected processing times and analysts' capacity constraints. Relative to discretionary allocation, the algorithmic approach is transparent, generates a more even distribution of workload across workers, and is optimized to yield a large number of completed cases. In contrast, discretionary assignment solves the same objective function as the algorithm incorporating soft information about worker-case fit, such as sector familiarity, writing quality, and reliability under time pressure. Both selection methods aim at maximizing output and preventing expiration risk by assigning cases under imperfect information about worker capacity and case complexity. The hard automation treatment simultaneously removes managerial judgment from case matching, and imposes workload equalization. We try to disentangle these two components in the mechanism analysis.

The assignment algorithm was constructed by OEFA with support from an external consultant. It solved a linear optimization problem ("knapsack problem") that maximizes the agency's productivity conditional on the capacity constraints of analysts, and the expected processing times for each task (the three enforcement stages) as a function of the cases' pre-analyzed complexity. These parameters were established by the agency. In practice, the algorithm simulated hundreds of allocations, and OEFA selected the one with the highest projected productivity, conditional on not having many cases at expiration risk. Appendix B provides a detailed description of each phase of the algorithm.

OEFA allocated all cases for analysis in the second semester of 2023, with analyses beginning in August 2023 and ending in December 2023. The goal was to distribute 1,415 cases to 104 analysts across 8 economic sectors (agriculture, commerce, electricity, oil and gas, industry, solid waste, mining, and fishing). OEFA ran the algorithm and selected the optimal allocation on June 2023, though 34 of them had not yet been hired at the time of the assignment so their identities were not known. Those lists were then given to the newly hired analysts according to the experimental protocol explained below. The algorithm filled each analyst's list with cases, and added suggested milestone dates for the three procedural documents.

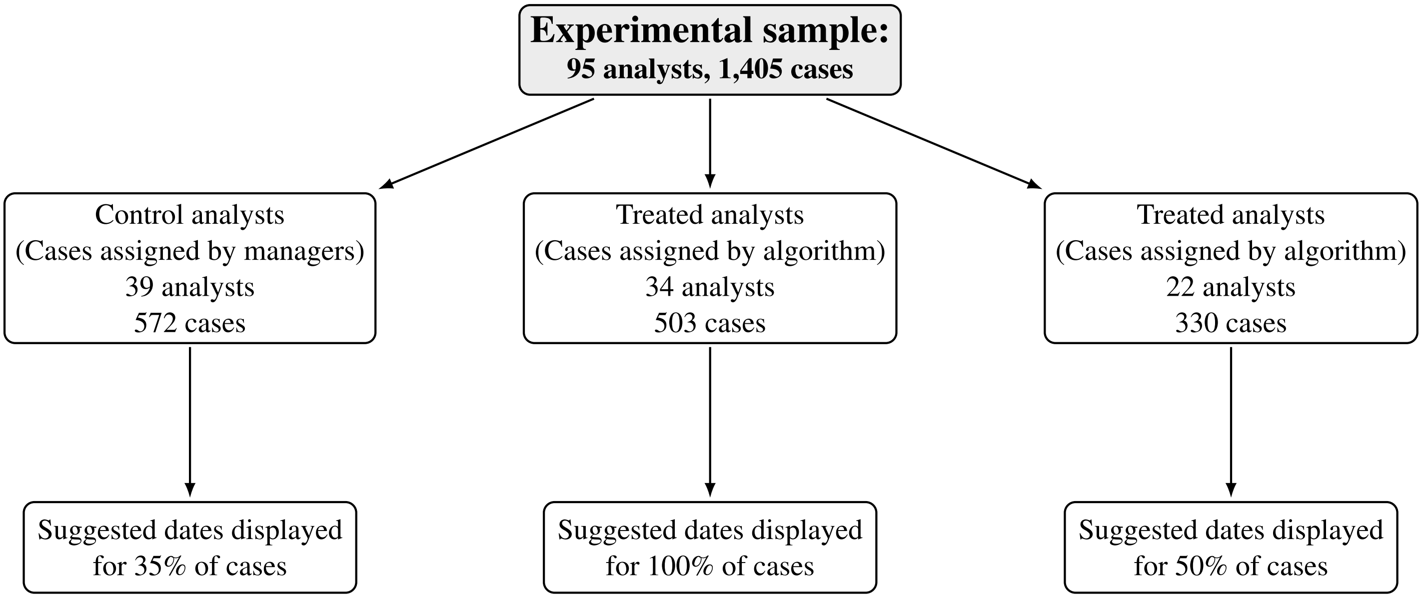
The experimental design divided analysts into a control group and a treatment group. In the final study sample, the control group contained 39 analysts and 572 cases, whereas the treatment group had 56 analysts and 833 cases. The treatment status was randomized at the analyst level stratified by eight economic sectors. Treated analysts received the algorithm-generated lists, while control analysts had their cases assigned by managers. Crucially, the algorithm first generated complete assignment lists for all analysts (treatment and control alike) but the algorithm-generated assignments for control-group analysts were then erased before distribution, and managers were asked to allocate those cases to control analysts at their discretion. This procedure ensures that both groups drew from the same pool of cases, isolating the matching component of managerial decisions from any confound arising from differences in case composition across groups. The cases were distributed to both groups between August and September 2023, and case processing was tracked through December 2023. At the time of randomization, many lists did not have an assigned analyst, since the new hires had not yet been assigned. The unidentified analysts corresponding to the new hires were evenly split between the groups. All lists were delivered in August 2023 to the analysts through Excel spreadsheets, following the usual practice inside the organization.

Besides assigning cases to workers, the algorithm also suggested completion dates. This additional information could act as a nudge by helping analysts organize their work, regardless of how their cases were assigned. To assess the role of these dates, we randomized their display at the case level for 35% of control-group cases and for 50% of the cases handled by 22 randomly selected treated analysts. The remaining 34 treated analysts saw suggested dates for all of their cases. Figure 1 illustrates the experimental design with the two levels of randomization: at the analyst level, and then at the case level. The randomization of suggested dates was only done with a sample of treated analysts because of a concern that it might weaken the treatment effect of algorithmic assignment. All suggested dates for conclusion lied before December 31, 2023, reflecting OEFA's intention to execute all the assigned cases in the same year.

In summary, as illustrated in Figure 1, the experiment generated three relevant groups: i) a control group of analysts whose cases were assigned by managers, with suggested dates displayed for 35% of cases; ii) a treatment group of analysts whose cases were assigned by the algorithm and for whom all cases displayed suggested dates; and iii) a treatment group of analysts whose cases were assigned by the algorithm but for whom suggested dates were displayed for only 50% of cases. The experiment was pre-registered in AsPredicted under 156452. The pre-registered hypotheses are stated in Appendix A. Cases were distributed to workers between August and September 2023, and case processing was tracked through December 2023.

There were a few noteworthy deviations from the research protocol. First, at the initial random assignment, in July 2023, there were 104 lists of 1,415 cases, but OEFA excluded a few lists and added a few other cases before distributing the lists to analysts. The exclusions resulted in a reduction of 90 cases, and 80 other cases were added (41 to control lists, and 39 to treatment lsits). As a result, we obtained the aforementioned sample of 1,405 cases and 95 analysts. These exclusions were mostly outright cancellations of full lists, meaning the full list was not found in the final assignment of cases, as opposed to piecemeal reductions in a

Figure 1: Experimental design with two cross-randomized treatments



Note: The figure illustrates the experimental design. Analysts were randomly assigned to a control group, whose cases were assigned by managers, and a treatment group, whose cases were assigned by the algorithm. The figure reports the number of analysts and cases in each experimental branch. Suggested completion dates were displayed for 35% of control-group cases, for all cases handled by 34 treated analysts, and for 50% of cases handled by 22 randomly selected treated analysts.

worker's load. Appendix Table B1 provides a detailed account of these deviations.

The exclusion of lists were balanced across control and treatment in the sectors agriculture, industry, solid waste, and fishing, while unbalanced in fuel retail, electricity, oil and gas, and mining. In general, there were more exclusions in the control group (a reduction from 46 to 39 lists) than in the treatment group (a reduction from 58 to 56 lists). To verify the importance of these deviations, we check for the balance of covariates across treatment and control groups.

Another deviation in the protocol was the reshuffling of cases across analysts, which altered the initial assignment. Of the 1,325 originally assigned cases (after the 90 exclusions), around 6% were reassigned to different analysts. Around 3% (45 cases) were initially assigned to unidentified control lists but then redistributed to treatment lists. The other reassigned cases were redistributed across treatment analysts. All things considered, the experiment was executed without major deviations from the research protocol.

One structural feature of the design merits attention as a potential limitation. By removing treatment analysts from managers' assignment pool, the experiment reduced the number of analysts for whom managers needed to make discretionary decisions. With a smaller group to oversee, managers may have been able to invest more attention per analyst in optimizing assignments—for example, by more carefully considering current workload, recent performance, or timing constraints. If the quality of managerial assignment improves as the scale of discretionary oversight decreases, the experiment may overstate the advantage of discretion relative to a scenario in which all analysts were under managerial assignment.

We also note a small deviation from the pre-analysis plan. While we follow closely the pre-analysis plan in testing the effects of hard and soft automation on timely case completion (hypotheses H1 and H2 in Appendix A), and quality of analysis (hypotheses H3 and H4), one of the pre-registered quality metrics could not be measured: the number of corrections made on the enforcement analysis. We use the other three quality metrics considered: the value of the fine imposed, the probability that the sanction is disputed by the regulated entity, the probability that the sanction is reverted on appeal.

5 Empirical Framework

The empirical analysis is conducted at the case level. To estimate the effect of hard automation—the replacement of managerial discretion in case assignment with algorithmic assignment—we estimate:

$$y_{ia} = \alpha + \beta_1 T_a + \beta_2 \tau_{ia} + \gamma X_{ia} + \varepsilon_{ia} \tag{1}$$

where y_{ia} is the outcome for case i and analyst a, T_a is the binary treatment assignment at the analyst level, τ_{ia} is the binary treatment of the suggested dates for case i and analyst a, and X_{ia} is a vector of case- and analyst-level characteristics: case complexity, analyst age, experience at OEFA, and whether the analyst worked remotely. ε_{ia} is an idiosyncratic shock. In Equation 1, the coefficients β_1 and β_2 capture the effect of automated assignment and suggested dates on the outcomes of interest: timely case completion, case expiration, and processing time.

Our main specification uses a binary treatment variable based on the initial assignment of analysts, which we call intent-to-treat effect (ITT). As mentioned earlier, deviations from the research protocol implied that some cases assigned to the control group were reassigned to treatment analysts, and some cases assigned to treatment analysts were reassigned to other treatment analysts. To address this non-compliance, we compute a continuous compliance measure containing the share of cases in the treatment group that were initially assigned as treatment. We then compute the Wald estimator through a 2SLS approach, using the initial random assignment as an instrument for actual assignment. We also report the OLS results showing the correlation between the continuous treatment variable and the outcomes.

Equation 1 estimates the effect of the soft automation (suggested dates) by using within and between-analyst variation. Equation 2 is used to estimate the effect of soft automation within analysts, since the treatment was randomized at the case level within the lists. We estimate:

$$y_{ia} = \delta + \omega\tau_i + \mu X_{ia} + \mu_a + \epsilon_{ia} \tag{2}$$

where τ_i is the binary treatment assignment at the case level, denoting whether the case displayed the suggested completion dates produced by the algorithm. μ_a are analyst fixed effects, included to isolate the effect of suggested dates net of time-invariant analyst ability. Many analysts' lists had no variation in τ_i and are therefore not used in the estimation of Equation 2. We study the same outcome family as in Equation 1.

In addition, we test whether analysts followed the suggested dates when organizing the order of their cases. To do so, we compute the difference between the actual and suggested dates, including for cases that did not display the suggested dates. If analysts used the scheduling information, realized start and completion dates should be closer to the algorithm's suggested dates among cases for which those dates were displayed.

In specifications that include controls, we add case- and analyst-level covariates to improve precision and help absorb residual imbalance, although identification does not rely on them given randomized assignment. Analyst-level controls include a dummy for "new employees" (employees who joined DFAI after January 1, 2023), age-tercile dummies, a master's-degree dummy, a high-experience dummy, and the analyst's caseload. Case-level controls comprise case complexity and assignment-month fixed effects. Because background information is unavailable for a small subset of analysts, we additionally include an indicator for missing analyst covariates, which allows us to preserve those analysts' cases in the estimation sample rather than exclude them from analysis.

In all specifications shown in the main results, standard errors are clustered at the analyst level. We also use perform inference using a randomization inference approach, as common in the experimental literature (Young 2019). Appendix E reports a randomization inference exercise that permutes treatment assignment within sectors across 1,000 draws and computes permutation-based p-values.

6 Data and Descriptive Statistics

6.1 Datasets Used

All data stems from administrative datasets from OEFA. The first dataset is the Stock of Cases (*expedientes*), which contains detailed information on all sanctioning procedures initiated during the experimental period. Each observation corresponds to a case, identified by a regulated entity and the type of alleged environmental infraction. The dataset includes information on the economic sector of the regulated entity, the assessed complexity of the case, and the unique identifiers for the three key procedural documents within OEFA's sanctioning mechanism: the Sub-Directorial Resolution (RSD), the Final Instruction Report (IFI), and the Directorial Resolution (RD). For each document, the database records the issue date and the time taken by the assigned analyst to complete the review. This dataset serves as the primary source for measuring case processing times and assessing the impact of the algorithmic assignment on procedural efficiency.

We also consider administrative data on analysts, such as their age, academic degree, professional experience (both general and specific to environmental enforcement), sectoral specialization, team leader, and date of hire. The high experience dummy used throughout the analysis is an indicator for analysts with above-median years of specific experience in legal analysis; the sample median is 5 years. These variables allow for the examination of potential heterogeneity in treatment effects and ensure that randomization achieved balance across analyst characteristics. Together, these datasets provide a comprehensive view of both the case management process and the personnel involved, enabling a rigorous evaluation of the algorithm's impact on the efficiency and consistency of environmental enforcement procedures within OEFA.

A third dataset, the Registry of Sanctioned Environmental Violators (RUIAS, for its acronym in Spanish), complements the analysis by tracking the downstream evolution of cases after the issuance of the RD. Unlike the internal administrative files, the RUIAS data were scraped from PIFA, OEFA's public online enforcement portal. For each case that resulted in a sanction, RUIAS records the final fine imposed, whether the regulated entity filed a reconsideration request or appealed to the Environmental Enforcement Tribunal (TFA, for its acronym in Spanish), and the outcome of those procedures. We refer to this as the Appeals dataset. These variables allow us to assess whether the two automation treatments affected the quality of enforcement decisions—specifically, whether cases processed under algorithmic assignment or with suggested dates produced sanctions that were more or less likely to be challenged or overturned on appeal.

6.2 Balance Tests

Table 1 reports balance tests comparing cases and analysts assigned to the discretionary and algorithmic groups. The table shows that randomization achieved close balance across all observed characteristics. At the case level, the share of complex cases is very similar across groups (0.488 in control versus 0.457 in treatment , p = 0.263), and the share of cases at risk of prescription is identical (0.149 in both groups, p = 0.989). These results indicate that the algorithmic and discretionary groups faced similar distributions of case difficulty and urgency.

Balance is also preserved at the analyst level. The month of assignment, the number of cases per analyst, and analyst demographics are statistically indistinguishable between groups. Analysts assigned to the algorithmic group are similar in age to those in the control group (32.3 versus 32.8 years, p = 0.806) and have comparable case loads (14.9 versus 14.7 cases, p = 0.857). Gender composition and educational attainment are also balanced, with no statistically significant differences in the share of female analysts or those holding a master's degree. The proportion of newly hired employees is nearly identical across treatment and control (0.643 versus 0.615, p = 0.788). This ensures that any differential performance of new versus experienced workers across assignment regimes cannot be attributed to compositional differences created by the randomization.

Overall, the balance tests confirm that the experimental design successfully equalized observed characteristics across the two assignment regimes. Consequently, differences in outcomes between algorithmic and discretionary assignment can be interpreted as causal effects of the assignment mechanism rather than artifacts of pre-existing differences in case mix or analyst composition.

Table 1: Balance Table

Variable	Control	Treatment	P-value
Complex Case	0.488	0.457	0.263
	(0.500)	(0.498)	
	[N=572]	[N=833]	
Risk of Prescription	0.149	0.149	0.989
	(0.356)	(0.356)	
	[N=572]	[N=833]	
Month Assignment	8.436	8.286	0.405
	(0.968)	(0.780)	
	[N=39]	[N=56]	
N. Cases	14.667	14.875	0.857
	(5.658)	(5.424)	
	[N=39]	[N=56]	
Age	32.821	32.339	0.806
	(9.580)	(9.225)	
	[N=39]	[N=56]	
Female	0.462	0.607	0.164
	(0.505)	(0.493)	
	[N=39]	[N=56]	
Masters' Degree	0.077	0.161	0.231
	(0.270)	(0.371)	
	[N=39]	[N=56]	
New Employee	0.615	0.643	0.788
	(0.493)	(0.483)	
	[N=39]	[N=56]	

Note: This table shows the mean, standard deviation (in parentheses) and number of observation (in square brackets) for several fixed characteristics of cases or analysts. The column for Control provides the statistics for those analysts and their cases in which the cases were assigned at discretion. The column for Treatment provides statistics for analysts and cases assigned by the algorithm. The third column, P-value, shows the p-value of a means-equality t-test between the two groups.

7 Results

This section presents the effects of hard automation on timely completion, then examines the behavioral response to suggested dates, heterogeneity across analysts and cases, and sanction and appeals outcomes.

7.1 Main Results

Table 2 presents the main estimates for the effect of hard automation on timely case completion, while also including the case-level indicator for soft automation. The outcome equals one if the case was concluded before December 31, 2023. This cutoff reflects the natural endpoint of the experimental period: all cases in the study were assigned between August and September 2023, DFAI's explicit goal was to complete them within the second semester of that year, and all suggested dates generated by the algorithm fell within this window. As a robustness check, Figure 2 shows that the negative effect of algorithmic assignment is not sensitive to the choice of cutoff date.

Because some originally treated analysts did not ultimately keep their algorithm-generated lists, the table reports three models that answer related but distinct questions. Columns (1)–(2) present intent-to-treat (ITT) estimates based on original random assignment; this is the preferred causal estimate, as original assignment is clean and unaffected by any post-randomization changes. Columns (3)–(4) present OLS estimates based on actual assignment after list changes; these are informative but less clean if implementation changes were endogenous to case characteristics. Columns (5)–(6) report IV estimates that instrument realized algorithmic assignment with original assignment, recovering the treatment-on-the-treated effect for analysts who complied with the original protocol.

Across specifications, cases assigned through hard automation are less likely to be completed on time. In the ITT specification with controls (Column (2)), original assignment to the algorithmic arm reduces timely completion by approximately 5 percentage points relative to a control-group mean of 40.4 percent. The corresponding realized-assignment estimate in Column (4) is of similar magnitude, while the IV estimate in Column (6) is somewhat larger. The consistency in sign across specifications indicates that the main result is not driven by the way non-compliance is handled. Substantively, these estimates imply that replacing managerial assignment with the algorithm-generated lists lowered the share of cases completed within the study period, even though Section 6.2 showed close balance in observed case difficulty, urgency, and analyst characteristics across arms.

The same table also shows a different pattern for soft automation, captured by the display of suggested completion dates. Displaying suggested completion dates increases the probability of timely completion by about 4 percentage points across specifications. Appendix Table C1 reports all control-variable coefficients from the with-controls specifications. Thus, the two interventions move outcomes in opposite directions within the same empirical framework: replacing managerial selection with algorithmic assignment reduces timely completion, whereas adding non-binding scheduling information increases it. The first-stage F-statistic in the IV specification exceeds 2,400, indicating very strong compliance between original and realized algorithmic assignment.

Figure 2 complements Table 2 by tracing the estimated effect of algorithmic assignment over alternative calendar cutoff dates between September 2023 and March 2024, serving as a robustness test for the main specification. The figure plots the corresponding difference in the probability that a case had been completed by that same date. The figure shows that the negative gap in completion widens over time, becoming most pronounced around the end of the calendar year and early 2024. This pattern is consistent with a setting in which initial differences in prioritization accumulate as analysts work through their lists.

Cases displaying suggested dates are approximately 4 percentage points more likely to be executed on time in the within-analyst comparison, and start earlier relative to the algorithm's proposed schedule. This effect is concentrated among cases assigned by managers; Section 8.3 provides a detailed analysis of the behavioral mechanism, including a test for whether the gain reflects pure salience or a general productivity effect.

As mentioned above, we also assess the statistical significance of the estimated results using randomization inference, which is particularly relevant given the small number of analyst clusters. Appendix E reports the results of a randomization inference exercise that permutes treatment assignment within sectors across 1,000 draws and computes permutation-based p-values. The randomization inference results are consistent with the main findings based on cluster-robust standard errors, confirming the statistical significance of the estimated effects of both hard and soft automation on timely completion. In this exercise, we run permutations at the analyst level, stratified by economic sector, which is consistent with the main design. The p-value of the two-sided test for the ITT effect is 0.07.

We also evaluate the heterogeneity of the treatment effects by the characteristics of analysts. Appendix

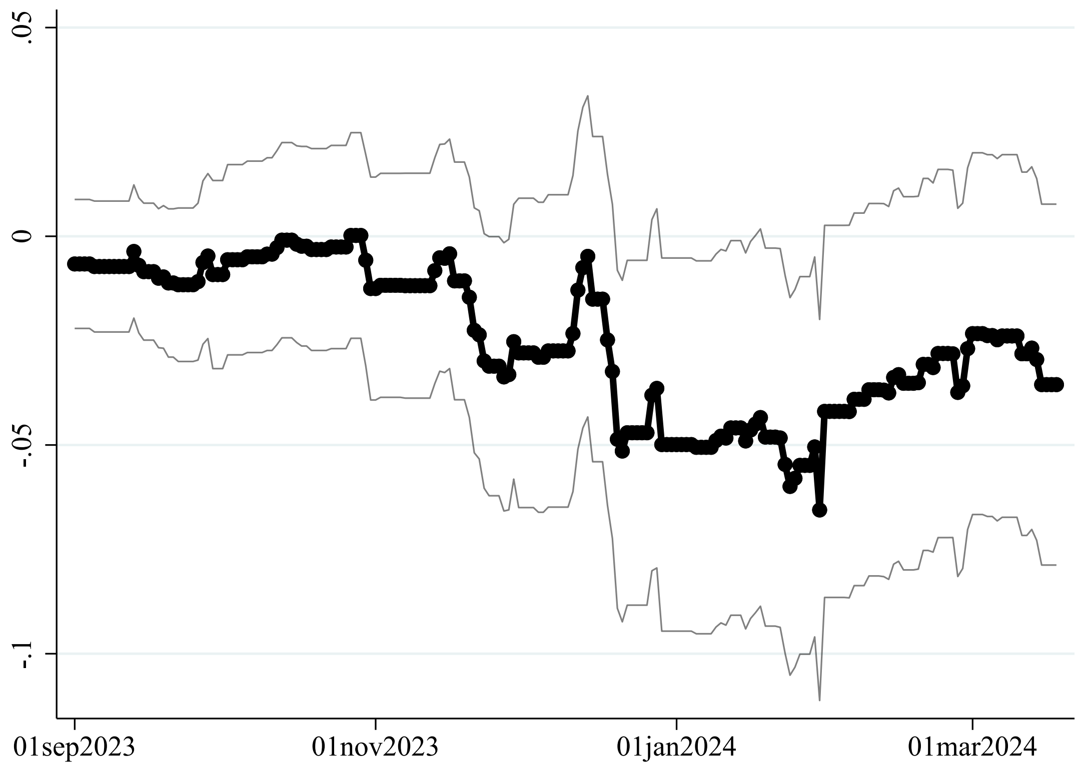
Table F1 shows the result of the hard automation treatment interacted with analyst experience and case complexity. Though the underlying characteristics are correlated with case execution, the interaction terms are never statistically significant, suggesting that the main effect of algorithmic assignment on timely completion is not driven by a particular subgroup of analysts or cases.

Table 2: Main Results – Probability of Concluding Case on Time

	ITT		OLS		IV	
	(1)	(2)	(3)	(4)	(5)	(6)
Original Algorithm Assignment	-0.043	-0.050*				
	(0.045)	(0.026)				
Actual Algorithm Assignment			-0.053	-0.056*	-0.048	-0.055*
			(0.045)	(0.028)	(0.049)	(0.028)
Case with Suggested Date	0.039	0.044*	0.042	0.045*	0.040	0.045*
	(0.032)	(0.024)	(0.031)	(0.024)	(0.032)	(0.023)
Controls	No	Yes	No	Yes	No	Yes
Mean Outcome	0.404	0.404	0.404	0.404	0.404	0.404
Mean Treatment	0.593	0.593	0.528	0.528	0.528	0.528
N. Inspectors	95	95	95	95	95	95
R2	0.184	0.303	0.184	0.303	0.184	0.303
N	1405	1405	1405	1405	1405	1405
First stage F					2721.762	3959.13

Notes: This table shows the results of the experiment on the probability that a case was concluded before December 31, 2023. Columns (1) and (2) show the Intention to Treat design, where the outcome (binary variable) is regressed on the original assignment of analysts into discretionary or algorithmic experimental arms. Columns (3) and (4) show the results for the same variable, where the algorithm assignment variable (at the analyst level) is "corrected" by the fact that OEFA effectively turned some treatment (i.e., algorithm assignment) analysts into discretionary assignment by substantially altering their lists. Columns (5) and (6) show results where this experimental non-compliance is dealt with a 2SLS approach, where the original assignment is used as an instrument for the actual assignment. Columns (1), (3), and (5) show results without any controls, but controlling for sector fixed effects (the randomization strata). Columns (2), (4), and (6) control for case characteristics (month of assignment and complexity) and analyst-level characteristics (new employee dummy, master's degree dummy, high experience dummy, dummies for age terciles, analyst caseload). Analyst-level characteristics are set to zero when they are missing, and we include a dummy for missing covariates. Standard errors are clustered at the analyst level. Significance levels: *10%, **5%, ***1%.

Figure 2: Coefficients on Algorithm Lists for Alternative Date Cut-Offs



Notes: This figure plots the estimated effect of algorithmic assignment on the probability that a case had been completed by each alternative date shown on the horizontal axis, ranging from September 2023 through March 2024. The estimates are based on OLS regressions of the outcome on the algorithmic assignment variable, controlling for sector fixed effects, case characteristics (month of assignment and complexity) and analyst-level characteristics (new employee dummy, master's degree dummy, high experience dummy, dummies for age terciles, analyst caseload). Analyst-level characteristics are set to zero when they are missing, and we include a dummy for missing covariates. Standard errors are clustered at the analyst level. The figure shows the point coefficient and a 90% confidence interval.

7.2 Secondary Outcomes

We next examine whether the two interventions affected other dimensions of case management and sanctioning outcomes. Table 3 examines whether the interventions affected other aspects of case organization and sanctioning. Columns (1) and (2) show no statistically significant effect of algorithmic assignment or suggested dates on the number of days between assignment and case start. Columns (3) and (4) similarly show no detectable effect on total processing days among completed cases. Columns (5) and (6) show no effect on the probability that the case ends with a penalty. Finally, Columns (7) and (8) examine fine amounts conditional on sanction. The coefficient on suggested dates is positive, around 0.28 log points, but imprecisely estimated in both specifications. Overall, the table indicates that the main effect of the interventions appears in timely completion rather than in these other dimensions of case handling. Because sanction and fine outcomes are only observed after case completion, these estimates are conditional on post-treatment completion and should be interpreted descriptively rather than as clean causal effects on unconditional enforcement quality.

We finally examine whether changes in completion and work sequencing are accompanied by changes in downstream appeals outcomes. Table 4 reports estimates of the effect of both automation treatments on enforcement quality, using data from the RUIAS dataset. The outcomes capture whether regulated entities exercised their right to appeal and whether the sanction changed as a result. Across Columns (1) and (2), there is no evidence that either intervention changes the probability that a regulated entity appeals. Columns (5) and (6) similarly show no clear effect on the number of appeals filed, and Columns (7) and (8) show no effect on the change in fine amounts after appeal. In Columns (3) and (4), suggested dates are associated with a higher number of infractions identified in the case, while algorithmic assignment itself has no detectable effect on that margin.

Because appeals outcomes are observed only after cases reach certain procedural stages or result in sanctions, conditioning on appeals-eligible cases may introduce post-treatment selection if the interventions affect completion. These outcomes should therefore be interpreted as evidence against large detectable quality deterioration among cases reaching the appeals stage, rather than as definitive evidence that assignment did not affect legal quality. Taken together, the appeals results indicate that hard automation did not improve downstream legal robustness, and provide only limited evidence that soft automation altered case content.

Table 3: Results on Work Organization and Sanction Outcomes

	Start Date		Days Process		P(Penalty)		log(Fine)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Original Algorithm Assignment	2.316		-2.771		-0.014		0.068	
	(2.408)		(3.163)		(0.031)		(0.192)	
Actual Algorithm Assignment		3.587		-4.068		-0.020		0.097
		(2.484)		(3.386)		(0.031)		(0.202)
Case with Suggested Date	-0.345	-0.795	-2.678	-2.225	-0.016	-0.014	0.284	0.274
	(1.842)	(1.864)	(4.093)	(4.132)	(0.034)	(0.034)	(0.174)	(0.175)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Outcome	39.025	39.025	86.476	86.476	0.868	0.868	2.053	2.053
Mean Treatment	0.590	0.590	0.556	0.556	0.556	0.556	0.551	0.551
N. Analysts	95	95	86	86	86	86	72	72
R2	0.101	0.102	0.110	0.111	0.112	0.112	0.346	0.346
N	1372	1372	567	567	567	567	396	396

Notes: This table shows the results of the experiments on secondary outcomes. Columns (1) and (2) use as outcome the number of days days between case assignment and case start. Columns (3) and (4) show the effect on the number of days between the end and start of the process. Columns (5) and (6) use as outcomes the probability that the case ended with a penalty. Columns (7) and (8) use the log of the fine. Columns (1), (3), (5), and (7) show results for the original algorithm assignment. Columns (2), (4), (6), and (8) show the results for the realized assignment after interference in the lists. All regressions control for sector fixed effects, case characteristics (month of assignment and complexity) and analyst-level characteristics (new employee dummy, master's degree dummy, high experience dummy, dummies for age terciles, analyst caseload). Analyst-level characteristics are set to zero when they are missing, and we include a dummy for missing covariates. Standard errors are clustered at the analyst level. Significance levels: *10%, **5%, ***1%.

Table 4: Results on the Outcomes of Case Appeals

	P(Appeal)		N. Infractions		N. Appeals		Change in Fine	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Original Algorithm Assignment	-0.000		-0.130		0.070		-0.028	
	(0.029)		(0.267)		(0.198)		(0.026)	
Actual Algorithm Assignment		0.004		-0.053		0.085		-0.032
		(0.030)		(0.283)		(0.203)		(0.027)
Case with Suggested Date	-0.006	-0.008	0.561**	0.526*	0.191	0.186	-0.055	-0.054
	(0.035)	(0.035)	(0.274)	(0.277)	(0.214)	(0.212)	(0.051)	(0.051)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Outcome	0.314	0.314	3.440	3.440	1.423	1.423	0.948	0.948
Mean Treatment	0.585	0.585	0.584	0.584	0.584	0.584	0.553	0.553
N. Analysts	95	95	95	95	95	95	90	90
R2	0.065	0.065	0.101	0.100	0.025	0.025	0.036	0.036
N	1167	1167	974	974	974	974	629	629

Note: This table reports estimates of the effect of algorithmic assignment and suggested dates on appeals-related outcomes drawn from the RUIAS dataset. Columns (1)–(2) show the effect on the probability that a regulated entity appeals the sanction. Columns (3)–(4) show the effect on the number of infractions identified in the case. Columns (5)–(6) show the effect on the number of appeals filed. Columns (7)–(8) show the effect on the change in the fine amount between the initial decision and the final ruling after appeals. Odd-numbered columns show intent-to-treat estimates using the original random assignment; even-numbered columns use the realized assignment. All regressions control for sector fixed effects, case characteristics (month of assignment and complexity), and analyst-level characteristics (new employee dummy, master's degree dummy, high experience dummy, dummies for age terciles, analyst caseload, missing-covariates indicator). Standard errors are clustered at the analyst level. Significance levels: *10%, **5%, ***1%.

8 Mechanisms

This section investigates the channels behind the main results. We test four hypotheses in turn. First, we examine whether the productivity cost of algorithmic assignment is concentrated among particular types of analysts or cases—specifically, whether newer workers or more complex cases drive the effect (Appendix F). Second, we test the *fairness mechanism*: if managers systematically overload some analysts, the algorithm's balancing constraint could improve efficiency by redistributing work, so we ask whether the treatment effect is larger where pre-existing workload inequality was greater (Section 8.1). Third, we test the *matching mechanism*: managers may use tacit knowledge about analyst skill to direct difficult cases toward able workers, so we examine whether this positive assortative matching is disrupted under the algorithm and whether the productivity gap is largest for complex assignments (Section 8.2). Fourth, we investigate the channel behind the suggested-dates effect: we ask whether scheduling cues raise absolute analyst productivity or operate purely through *salience*, making dated cases more prominent relative to undated ones on the same list (Section 8.3).

8.1 Workload Equalization and Fairness

One channel through which automated assignment could improve productivity is workload equalization. To the extent that unequal workloads lower morale and generate perceived unfairness, the equalization of workload through automated assignment might improve productivity, a potential mechanism suggested previously in the management literature (Bai et al. 2022; Pechová, Volfová, and Jírová 2023; Beer, Qi, and Rios 2025; Bartling, Fehr, and Schmidt 2013).

To test this, we compute three measures of unequal workload distribution (i.e., number of assigned cases) and interact them with the algorithm indicator: the standard deviation of case workload across analysts by sector (*SD sector*), the standard deviation of caseloads within team leaders (*SD manager*), and a dummy indicating that the standard deviation by team leader is greater than the median (*Unequal manager*). These measures are computed using the assignments as they were made during the experimental period, so they are also affected by the algorithm's equalization. However, they are able to capture the heterogeneity of unequal workload across sectors (the stratum of randomization) and team leaders, thus allowing us to test for this mechanisms through a heterogeneity analysis.

Table 5 interacts the algorithm indicator with three measures of the degree of workload inequality in the pre-period: the standard deviation of caseloads within sectors (*SD sector*), within team leaders (*SD manager*), and the binary indicator. Across all three specifications, the interaction terms are small and statistically indistinguishable from zero, suggesting that the algorithm's effect on timely completion does not vary with the degree of inequality in workload distribution in the team, defined by economic sector of team leader. The coefficients on the inequality are themselves small and statistically indistinguishable from zero, except for Column 3, where the teams whose managers have more unequal workload distributions seem to perform better in timely completion. However, there is no difference in the effect of the algorithm between these teams and those with more equal workload distributions. The main coefficients on algorithm assignment and case with suggested dates remain stable across these specifications.

Table 5: Fairness Mechanism: Heterogeneity by Pre-existing Workload Inequality

Outcome: Executed on Time	Inequality by sector	Inequality by manager	Dummy
	(1)	(2)	(3)
Algorithm Assignment	-0.086**	-0.087**	-0.085**
	(0.037)	(0.034)	(0.034)
Case with Suggested Date	0.048*	0.047*	0.046*
	(0.025)	(0.025)	(0.025)
Algorithm × SD (Sector)	0.006		
	(0.012)		
Algorithm × SD (Manager)		0.007	
		(0.012)	
SD Caseload (Manager)		0.015	
		(0.013)	
Algorithm × Unequal Manager			0.028
			(0.051)
Unequal Manager			0.170**
			(0.069)
Controls	Yes	Yes	Yes
Mean Outcome	0.404	0.404	0.404
N. Inspectors	95	95	95
R2	0.290	0.294	0.299
N	1405	1405	1405

Note: This table shows results of the experiment on timely case completion, interacting algorithmic assignment with measures of workload dispersion under discretionary assignment. Column (1) interacts with the standard deviation of caseloads within the sector; Column (2) with the standard deviation within team leaders; Column (3) with a binary indicator for team leaders with above-median caseload dispersion. All columns include sector fixed effects and controls for case characteristics (month of assignment, complexity) and analyst-level characteristics (new employee dummy, master's degree dummy, high experience dummy, age-tercile dummies, analyst caseload, missing-covariates indicator). Standard errors are clustered at the analyst level. Significance levels: *10%, **5%, ***1%.

8.2 Matching Complex Cases to Experienced Analysts

A second candidate mechanism is matching quality: managers may use soft knowledge about analyst skill to direct complex cases toward more experienced workers, thereby reducing the risk of errors or delays on difficult assignments. This hypothesis has been suggested in previous literature that documented positive effects of assignment discretion on bureaucrat performance (Duflo et al. 2018, Bachas et al. 2025).

To test this mechanism, we leverage the pre-processed case complexity variable, which is a binary indicator of whether the case was deemed complex by OEFA's internal standards. This classification was made before the randomization, and the distribution of complex cases was similar across the control and algorithm groups. We first test whether the assignment of complex cases is correlated with analyst quality, measured by education, experience in legal analysis, and time at OEFA. We investigate how this correlation differs between the two assignment regimes. We then test whether the productivity cost of algorithmic assignment is larger for complex cases.

Panel A of Table 6 shows the correlation between complex case assignment and a dummy for experienced analysts, conditional on several other controls. Column (1) regresses case complexity on analyst characteristics within the control group, while Column (2) repeats the exercise for the algorithm group. Column (3) does the same exercise for the whole sample, adding interaction terms for the characteristics of analysts and the treatment status. While this table only shows the coefficient for the variable experience, Appendix Table F2 shows the coefficients for the other quality variables, which are not significant. These results strongly suggest that, in the control group, more experienced analysts were more likely to be assigned complex cases, consistent with managers using their discretion to match difficult cases to more able workers. In contrast, this pattern is attenuated in the algorithm group, which is consistent with the algorithm's equalization objective disrupting this positive assortative matching.

Panel B of the same table tests whether the productivity loss in algorithmic assignment is being driven by these complex cases, which are no longer being assigned to experienced analysts. Again, Columns (1), (2), and (3) refer to the control, algorithm, and whole sample with interactions, respectively. The coefficient on complex cases is not significant for the control group, but it is negative and significant for the algorithm group. We can also see, in Column (3) that the standalone coefficient on the algorithm assignment becomes small and insignificant after controlling for the interaction of algorithm and complex cases, suggesting that the main driver of the loss in productivity comes from this set of cases.

Table 6: Matching Mechanism: Complexity Assignment and Execution Under Algorithm

Panel A: P(Complex Case)

	(1)	(2)	(3)
Experienced Worker	0.153***	0.030	0.139***
	(0.041)	(0.036)	(0.041)
Experienced W. x Algorithm			-0.114**
			(0.055)
Sample	Control	Algorithm	All
Mean Outcome	0.488	0.457	0.470
R2	0.281	0.192	0.222

Panel B: P(Timely Completion)

	(1)	(2)	(3)
Complex Case	-0.046	-0.083**	-0.013
	(0.056)	(0.032)	(0.045)
Algorithm Assignment			-0.008
			(0.034)
Algorithm $\times$ Complex Case			-0.087*
			(0.051)
Sample	Control	Algorithm	All
Mean Outcome	0.441	0.378	0.404
R2	0.343	0.295	0.305
N. Inspectors	39	56	95
N. Cases	572	833	1405

Panel A of Table 6 examines whether experienced analysts were more likely to receive complex cases under discretionary versus algorithmic assignment. The outcome is a binary indicator for whether the case is complex. Panel B examines whether the effect of algorithmic assignment on timely completion differs by case complexity. Columns (1) and (2) restrict the sample to control and algorithm-assigned analysts, respectively; Column (3) pools both groups and includes the relevant interaction term. All specifications include sector fixed effects and the controls described in the text. Standard errors are clustered at the analyst level. Significance levels: *10%, **5%, ***1%.

8.3 Suggested Dates: Salience vs Productivity

To unpack the impact of suggested dates, we examine the within-analyst variation in the effect of suggested dates on case completion more closely. Since the randomization was done at the case level, we can run the regressions in Equation 2 to estimate the effect of suggested dates on timely completion, controlling for analyst fixed effects. This exercise allows us to isolate the effect of suggested dates from any differences in analyst ability or motivation, since we are comparing cases handled by the same analyst with and without suggested dates. For this exercise, we exclude analysts in the treatment group who had no variation in the display of suggested dates, since they do not contribute to the estimation of the within-analyst effect.

Table 7 shows that cases displaying suggested dates are approximately 4 percentage points more likely to be executed on time in the within-analyst comparison, and start earlier relative to the algorithm's proposed schedule. The table also shows that this effect is concentrated among cases assigned by managers.

We investigate whether the positive effect of suggested dates reflects a real productivity gain or mere salience, analysts prioritizing dated cases at the expense of undated ones on the same list without raising overall output. Two sources of variation allow us to test this. Among algorithm-assigned analysts, random assignment to an all-dates group (100% of cases with suggested dates) versus a partial-dates group (50%) provides quasi-random variation in the share of dated cases across analysts. Among control analysts, managers assigned dated cases in varying quantities, generating natural cross-analyst variation in that share.

Table 7 presents the results in two panels. Panel A estimates the within-analyst effect of suggested dates on timely completion using analyst fixed effects, separately for control and treated analysts. The effect is positive and significant for control analysts, but smaller and imprecisely estimated for treated analysts with partial date variation; the interaction term in Column (4) confirms that the within-analyst benefit is concentrated among discretionary-assignment cases. Panel B tests for general productivity effects at the analyst level by asking whether analysts with a higher share of dated cases complete more cases in total. The evidence is consistent with suggested dates raising completion partly through the salience of individual dated cases, though the analyst-level results do not rule out a broader productivity effect among control analysts.

Table 7: Soft Automation – Probability of Concluding Case on Time (Within-Analyst Variation)

Panel A: Analyst-level Fixed Effects

	(1)	(2)	(3)	(4)
Case with Suggested Date	0.053*	0.005	0.036	0.059**
	(0.030)	(0.052)	(0.027)	(0.029)
Case with Suggested Date x Algorithm				-0.055
				(0.060)
Sample	Control	Algorithm	All	All
Mean Outcome	0.441	0.378	0.404	0.404
Mean Treatment	0.343	0.815	0.623	0.623
R2	0.393	0.342	0.361	0.361

Panel B: Sector-level Fixed Effects

	(1)	(2)	(3)	(4)
Share Cases with Suggested Date	0.281**	0.027	-0.034	0.232***
	(0.110)	(0.065)	(0.041)	(0.080)
Share Cases x Algorithm				-0.195***
				(0.057)
Sample	Control	Algorithm	All	All
Mean Outcome	0.441	0.378	0.404	0.404
Mean Treatment	0.343	0.815	0.623	0.623
R2	0.333	0.281	0.289	0.293
N. Inspectors	39	56	95	95
N. Cases	572	833	1405	1405

Notes: Notes: This table examines the effect of suggested completion dates on the probability that a case was concluded before December 31, 2023. Panel A uses within-analyst variation in whether cases displayed suggested dates and includes analyst fixed effects, thereby comparing dated and undated cases handled by the same analyst. Panel B relates timely completion to the share of an analyst's cases that displayed suggested dates and includes sector fixed effects. Columns (1) and (2) report estimates separately for control and algorithm-assigned analysts, respectively. Columns (3) and (4) pool both groups, with column (4) including an interaction between suggested dates and algorithmic assignment. All specifications control for case complexity and month-of-assignment fixed effects. Standard errors are clustered at the analyst level. Significance levels: *10%, **5%, ***1%.

8.4 Discussion

The findings above are consistent with the interpretation that discretionary managers embed tacit knowledge about worker-case fit into their assignments, and that the algorithm's inability to replicate this matching contributes to the observed productivity loss. Although the algorithm equalized workloads across analysts, it did not account for differences in skill or experience that may be critical for handling complex cases efficiently. The disruption of matching between experienced analysts and complex cases appears to be a key driver of the productivity decline under algorithmic assignment. This mechanism is consistent with prior evidence from tax enforcement and environmental inspections, and highlights the importance of tacit knowledge in bureaucratic production processes.

The results also highlight a distinct role for light-touch process design. Providing suggested completion dates increases the likelihood that cases are completed on time and shifts work closer to planned schedules, consistent with models in which deadlines serve as cognitive anchors that help workers organize effort and manage attention. However, these benefits are only observed among analysts who had their cases assigned by team managers. This pattern reinforces the idea that the automation gains are realized in settings where it complements human discretion rather than replacing it.

Indeed, among analysts in the control group, managers could choose freely how to allocate the cases that had suggested dates attached to them, and they did so in unbalanced ways across analysts. It is possible, for example, that they assigned these cases to analysts who might benefit from the scheduling information, or that they used the suggested dates to structure the sequencing of cases in a way that improved performance. Our results suggest that these analysts not only were more productive on the dated cases relative to the undated cases, but were also more productive overall, suggesting that the scheduling information may have had a broader effect on their workflow and organization of effort.

9 Conclusion

This paper provides experimental evidence on the productivity effects of two forms of algorithmic automation in a large public-sector bureaucracy. Using a pre-registered field experiment at Peru's national environmental enforcement agency, we find that replacing managerial case assignment with an optimization algorithm reduced timely case completion by approximately 5 percentage points relative to a control mean of 40 percent. Providing analysts with suggested milestone dates, in contrast, increased timely completion by 4 percentage points. We find no evidence that either treatment affected appeal rates, appeal outcomes, or fine amounts, though suggested dates are associated with a higher number of infractions among observed sanctioning outcomes. The results are consistent with full-replacement algorithmic assignment reducing productivity when managers hold valuable soft information about worker-case fit, and with lighter-touch automation improving workflow without displacing discretion.

Our findings contribute to a growing literature on algorithmic management and bureaucratic discretion. Prior work in tax enforcement and environmental inspections has found that full-replacement random assignment can reduce performance when discretionary systems embed tacit knowledge (Bachas et al. 2025; Duflo et al. 2018). The paper extends this literature to a legally constrained enforcement setting with strict statutory deadlines, and identifies the disruption of positive assortative matching between experienced analysts and complex cases as a plausible mechanism. The contrast between hard and soft automation also adds to evidence on the merits of rules-based replacement versus light-touch process design in high-stakes bureaucratic settings.

Several questions remain open for future research. First, the experiment evaluates one specific full-replacement regime; hybrid designs that combine algorithmic recommendations with managerial override authority could yield different outcomes and warrant direct testing. Second, the study covers a six-month period of unusually high workload and temporary staffing expansion; longer-run effects on learning, morale, and analyst retention remain unknown. Third, the algorithm studied here was static and did not adapt to observed worker performance; learning-based systems that update assignments dynamically may mitigate the matching losses documented here.

References

Altmann, Steffen, Christian Traxler, and Philipp Weinschenk (2022). "Deadlines and memory limitations". *Management Science* 68.9, pp. 6733–6750.

Bachas, Pierre, Anne Brockmeyer, Alipio Ferreira, and Bassirou Sarr (2025). "Algorithms and bureaucrats: evidence from tax audit selection in Senegal". *The World Bank Working Paper Series*.

Bai, Bing, Hengchen Dai, Dennis J Zhang, Fuqiang Zhang, and Haoyuan Hu (2022). "The impacts of algorithmic work assignment on fairness perceptions and productivity: Evidence from field experiments". *Manufacturing & Service Operations Management* 24.6, pp. 3060–3078.

Banerjee, Abhijit, Esther Duflo, Clement Imbert, Santhosh Mathew, and Rohini Pande (2020). "E-governance, accountability, and leakage in public programs: Experimental evidence from a financial management reform in India". *American Economic Journal: Applied Economics* 12.4, pp. 39–72.

Bartling, Björn, Ernst Fehr, and Klaus M Schmidt (2013). "Discretion, productivity, and work satisfaction". *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, pp. 4–22.

Beer, Ruth, Anyan Qi, and Ignacio Rios (2025). "Behavioral externalities of process automation". *Management Science*.

Benlian, Alexander, Martin Wiener, W Alec Cram, Hanna Krasnova, Alexander Maedche, Mareike Möhlmann, Jan Recker, and Ulrich Remus (2022). "Algorithmic management: bright and dark sides, practical implications, and research opportunities". *Business & Information Systems Engineering* 64.6, pp. 825–839.

Campbell, Donald J and Karl F Gingrich (1986). "The interactive effects of task complexity and participation on task performance: A field experiment". *Organizational behavior and human decision processes* 38.2, pp. 162–180.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan (2018). "The value of regulatory discretion: Estimates from environmental inspections in India". *Econometrica* 86.6, pp. 2123–2160.

Finan, Frederico, Benjamin A Olken, and Rohini Pande (2017). "The personnel economics of the developing state". *Handbook of economic field experiments* 2, pp. 467–514.

Haseeb, Muhammad and Kate Vyborny (2022). "Data, discretion and institutional capacity: Evidence from cash transfers in Pakistan". *Journal of Public Economics* 206, p. 104535.

Kang, Karam and Bernardo S Silveira (2021). "Understanding disparities in punishment: Regulator preferences and expertise". *Journal of Political Economy* 129.10, pp. 2947–2992.

Kellogg, Katherine C, Melissa A Valentine, and Angele Christin (2020). "Algorithms at work: The new contested terrain of control". *Academy of management annals* 14.1, pp. 366–410.

Khan, Adnan, Asim Khwaja, and Benjamin A Olken (2016). "Tax farming redux: Experimental evidence on performance pay for tax collectors". *The Quarterly Journal of Economics* 131.1, pp. 219–271.

— (2019). "Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings". *American Economic Review* 109.1, pp. 237–270.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2018). "Human decisions and machine predictions". *The quarterly journal of economics* 133.1, pp. 237–293.

Knowles, Stephen, Maroš Servátka, Trudy Sullivan, and Murat Genç (2022). "Procrastination and the non-monotonic effect of deadlines on task completion". *Economic inquiry* 60.2, pp. 706–720.

Ludwig, Jens, Sendhil Mullainathan, and Ashesh Rambachan (2024). "The unreasonable effectiveness of algorithms". *AEA Papers and Proceedings*. Vol. 114. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 623–627.

Margalit, Yotam and Shir Raviv (2024). "When Your Boss is an Algorithm: The Effect of Algorithmic Management on Worker Performance". *Available at SSRN 4776355*.

Olken, Benjamin A (2006). “Corruption and the costs of redistribution: Micro evidence from Indonesia”. *Journal of public economics* 90.4-5, pp. 853–870.

— (2007). “Monitoring corruption: evidence from a field experiment in Indonesia”. *Journal of political Economy* 115.2, pp. 200–249.

Pape, Tom, Stelios Kavadias, and Svenja C Sommer (2024). “Decision bias in project selection: Experimental evidence from the knapsack problem”. *Available at SSRN 3448676*.

Pechová, Jana, Hana Volfová, and Anežka Jírová (2023). “Impact of task assignment on effectiveness in work teams”. *Entrepreneurship and Sustainability Issues* 10.4, p. 152.

Schiffels, S, T Fliedner, and R Kolisch (2013). “How Decision Makers Solve the Knapsack Problem: Insights from Experimental Studies”. *Working paper*.

Young, Alwyn (2019). “Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results”. *The quarterly journal of economics* 134.2, pp. 557–598.

Appendix

A Pre-Analysis Plan

The experiment is designed to test the following pre-registered hypotheses:

- H1: Algorithmic assignment increases the probability of timely case completion relative to discretionary assignment.
- H2: Suggested completion dates increase the probability of timely case completion relative to no scheduling cues.
- H3: Algorithmic assignment improves the quality of enforcement decisions, as measured by the value of fines imposed, the probability that sanctions are disputed by regulated entities, and the probability that sanctions are reverted on appeal.
- H4: Suggested completion dates do not affect the quality of enforcement decisions.

To clarify mechanisms, we also examine the following secondary hypotheses:

- H1a: The reduced variance in workloads under algorithmic assignment increases timely case completion.
- H2a: Discretionary assignment allows managers to leverage tacit information about worker-case fit, which increases timely case completion.

B Assignment Algorithm: Implementation Details

This appendix describes the Python-based algorithm used to assign enforcement cases (*expedientes*) to legal analysts at OEFA. The algorithm was implemented in 2023 in four sequential phases.

Phase 1: Random Assignment

The algorithm begins with a database of 1,415 enforcement cases. Each case is characterized by the sector it belongs to (Agriculture, Commerce, Infrastructure and Services, Electricity, Major Hydrocarbons, Industry, Mining, and Fisheries) and a complexity level (very high, high, medium, low, or very low). Complexity determines the legal processing time allotted to each case stage, which varies by sector.

Cases are first divided by sector and assigned a priority flag based on their prescription deadline: cases with deadlines falling before December 31, 2023 receive higher priority. Within each sector-priority cell, cases are randomly shuffled and then split as evenly as possible among the legal analysts in that sector's team. Within each analyst's queue, cases are sorted by proximity to their prescription deadline, so that those at greatest risk of expiring are placed first in line.

To ensure robustness, this random shuffling procedure is repeated 100 times using different random seeds, producing 100 distinct assignment scenarios.

Phase 2: Optimal Case Selection

For each of the 100 random assignment scenarios, the algorithm determines which cases each analyst should actually work on, given limited time. This is formulated as a *knapsack problem* and solved using the Gurobi mathematical optimization solver.

Each analyst faces two constraints: the total processing time of selected cases cannot exceed their individual time capacity, and the total number of selected cases cannot exceed a per-capita reviewer capacity. The objective is to maximize the total value of cases selected, where value reflects the importance or urgency of resolving each case. The optimization is run over two sequential time periods (months). Cases not selected in the first period are carried forward to the second, but with a reduced capacity (one-third less), reflecting that the analyst has less remaining time.

The output of this phase is a binary indicator for each case: selected (assigned to the analyst's active workload) or not selected (deferred or unassigned).

Phase 3: Scheduling

Once the optimal subset of cases is identified for each analyst, the algorithm assigns specific start and end dates to each case stage. Each case passes through three procedural stages: the Sub-Directorate Resolution (RSD), the Final Instruction Report (IFI), and the Directorate Resolution (RD). Planned durations for each stage are determined by the case's sector and complexity.

Dates are computed in business days, excluding weekends and official Peruvian public holidays. Each analyst's cases are split into two cohorts. The first case in each cohort begins at the designated start of the corresponding procedural window; subsequent cases begin one business day after the previous case concludes.

Two scheduling variants are implemented. Under *quota-based scheduling*, the two cohorts are scheduled independently, each starting from the beginning of its respective time window. Under *sequential scheduling*, the second cohort begins where the first cohort left off, creating a continuous chain.

Phase 4: Evaluation and Seed Selection

The final phase evaluates all 100 random-seed scenarios and identifies which produces the best outcome. The criterion is the number of cases with a scheduled completion date on or before December 31, 2023. The algorithm identifies cases that appear in the assigned workload but are not present in the original case database (i.e., cases that could not be accommodated), and produces a graphical and tabular summary for each seed. The seed that maximizes the number of cases completed within the calendar year is selected as the final assignment.

Summary

In practice, the algorithm operates as follows. Starting from the full pool of cases, it randomly shuffles cases across 100 scenarios and distributes them equally among analysts. For each scenario, it solves an optimization problem to select the feasible subset of cases that maximizes total value subject to time and reviewer capacity constraints. It then constructs a day-by-day schedule for each analyst's selected cases, respecting legal stage durations and official holidays. Finally, it picks the scenario that results in the most cases being closed before the year's end. This final scenario constitutes the treatment arm's actual case assignments in the experiment.

Sample Construction

Table B1 traces the transition from the algorithm's initial case pool to the final analysis sample.

Table B1: Sample Construction and Implementation Deviations

Step	Cases	Analysts	Note
Experimental assignment generated	1,415	104	Original intended experiment
Cases/lists removed before implementation	90	9	
Replacement cases added	80		
Final analysis sample	1,405	95	Main estimation sample

Note: This table documents the transition from the algorithm's initial input pool to the final analysis sample used in estimation.

Table B2: Summary of Cases and Analysts in the Experiment

Sector	N. Cases Discretion (1)	N. Cases Algorithm (2)	N. Analysts Control (3)	N. Analysts Algorithm (4)	% Complex Cases (5)	% Cases with Prescription Risk (6)
Agriculture	29	28	2	2	58	0
Fuel Retail	134	290	6	12	20	22
Electricity	45	57	3	4	61	5
Oil and Gas	68	122	7	12	79	29
Industry	118	118	8	8	54	2
Solid Waste	30	30	2	2	43	0
Mining	88	128	8	12	66	24
Fishing	60	60	4	4	28	0
Total	572	833	39	56	47	15

Notes: This table summarizes key numbers in the algorithm, separated by economic sector (rows), which was the level of stratified randomization. Column (1) shows the number of enforcement cases assigned at discretion by the team leaders, and Column (2) the number of algorithmically-assigned cases. Columns (3) and (4) show the number of analysts that had their cases assigned at discretion and by the algorithm, respectively. Column (5) shows the proportion of complex cases within each sector. Column (6) shows the percentage of cases that were at risk of prescription, meaning that their prescription date was between August and December 2023.

C Main Results: Full Coefficient Table

Table C1 replicates the with-controls columns of Table 2 for each of the three specifications (ITT, OLS, IV) but reports all coefficient estimates, including the analyst- and case-level controls. The table confirms that the treatment effects are not driven by any single control variable and that the estimated magnitudes are consistent across specifications. Complex cases are less likely to be completed on time, while analyst caseload is positively associated with timely completion. New-employee status and experience are not statistically significant in these specifications. Case complexity is negatively associated with timely completion in most specifications. The stability of the treatment coefficient across columns with and without controls (Table 2 versus Table C1) reflects the success of the randomization in achieving balance on observed characteristics.

Table C1: Main Results with All Control Coefficients

	(ITT)	(OLS)	(IV)
Original Algorithm Assignment	-0.050*		
	(0.026)		
Case with Suggested Date	0.044*	0.045*	0.045*
	(0.024)	(0.024)	(0.023)
New Employee	-0.041	-0.041	-0.041
	(0.036)	(0.036)	(0.035)
Master's Degree	-0.009	-0.009	-0.009
	(0.033)	(0.033)	(0.033)
Experienced Worker	-0.018	-0.017	-0.017
	(0.032)	(0.032)	(0.031)
Complex Case	-0.064**	-0.064**	-0.064**
	(0.029)	(0.029)	(0.028)
N. Cases (Analyst)	0.011***	0.011***	0.011***
	(0.002)	(0.002)	(0.002)
Actual Algorithm Assignment		-0.056*	-0.055*
		(0.028)	(0.028)
Controls	Yes	Yes	Yes
Mean Outcome	0.404	0.404	0.404
Mean Treatment	0.593	0.528	0.528
N. Inspectors	95	95	95
R2	0.303	0.303	0.303
N	1405	1405	1405
First stage F			3959.13

Note: This table reproduces the with-controls specifications from Table 2 and reports all coefficient estimates, including analyst- and case-level controls. Column (ITT) uses the original experimental assignment (treatment_binary) as the regressor; Column (OLS) uses the realized assignment share (treatment_continuous); Column (IV) instruments the realized share with the original assignment. Controls include analyst age terciles, new-employee indicator, master's degree indicator, high experience indicator, complex-case indicator, missing-covariates indicator, analyst caseload, and month-of-assignment fixed effects. Standard errors are clustered at the analyst level. Significance levels: *10%, **5%, ***1%.

D Results on Start Date

Figure D1 plots the difference between algorithmic and discretionary lists in the probability that a case had been started by each date. The estimates are small, and non-significant throughout the period, indicating that cases on algorithmic lists were no less likely to have been started by any given date.

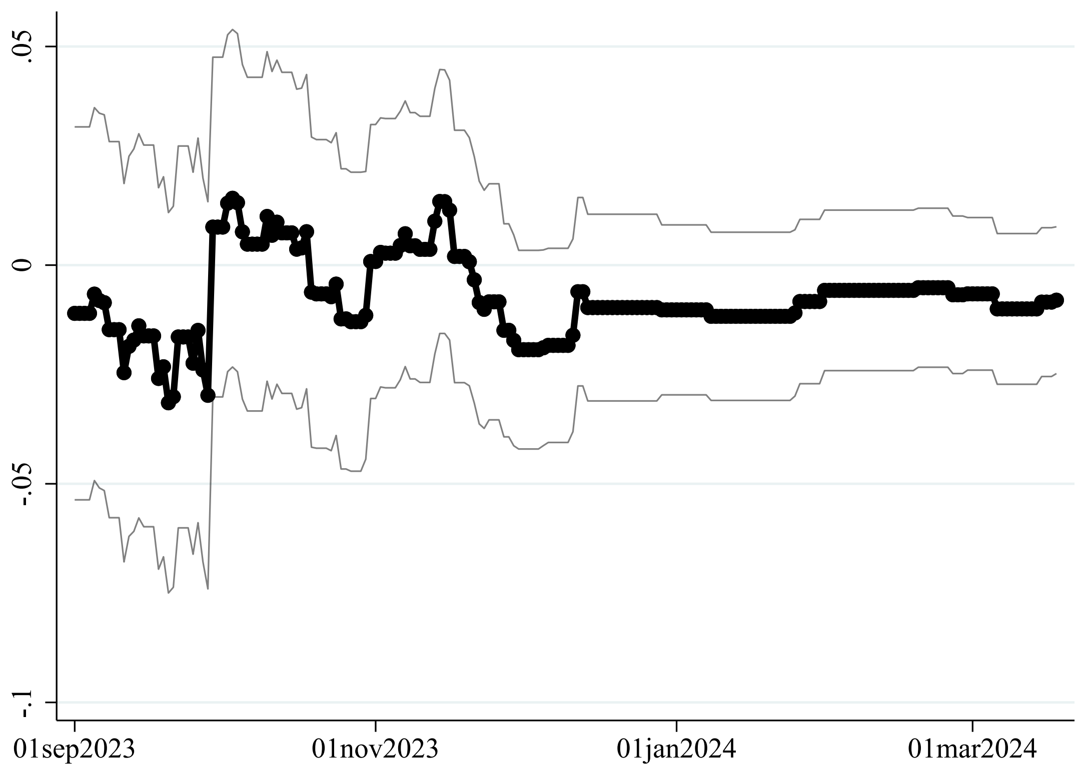


Figure D1: Difference in P(Start) by Date

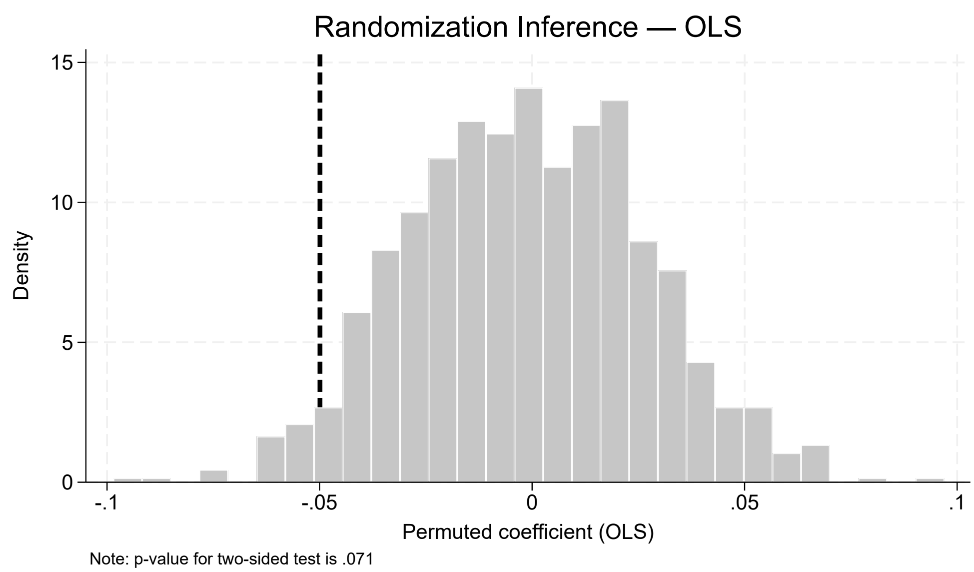
Notes: This figure plots the estimated effect of algorithmic assignment on the probability that a case had been started by a given date, as indicated by the date of the first document of the administrative process. The estimates are based on OLS regressions of the outcome on the algorithmic assignment variable, controlling for sector fixed effects, case characteristics (month of assignment and complexity) and analyst-level characteristics (new employee dummy, master's degree dummy, high experience dummy, dummies for age terciles, analyst caseload). Analyst-level characteristics are set to zero when they are missing, and we include a dummy for missing covariates. Standard errors are clustered at the analyst level. The figure shows the point coefficient and a 90% confidence interval.

E Randomization Inference

To assess whether the main treatment effect is robust to the small number of clusters (95 analysts), we implement a randomization inference procedure. We approximate the distribution of coefficients by permuting the treatment vector across analysts, preserving the stratified structure of the randomization (i.e., permuting within sectors), and re-estimating the main OLS regression for each of 1,000 permuted treatment assignments.

The proportion of permuted coefficients whose absolute value exceeds the absolute value of the observed coefficient gives the p-value for a two-sided test of the null hypothesis. Figure E1 shows the resulting permutation distributions for the OLS specifications, with the observed coefficient marked as a vertical dashed line. The observed coefficient lies in the left tail of the distribution. The results corroborate the conclusion that the estimated negative effect of algorithmic assignment on timely completion is unlikely to be a chance finding. The estimated coefficient lies roughly at the 3rd percentile of the permutation distribution, corresponding to a randomization inference p-value of approximately 0.07 for a two-sided test.

Figure E1: Randomization Inference: Permutation Distributions



Note: This figure shows the distribution of the ITT coefficient on algorithmic assignment across 1,000 permutations of the treatment vector, permuting analysts within sectors (the stratum of randomization in the study). The vertical line marks the observed coefficient from the main specification.

F Heterogeneity Analysis

Table F1 shows the result of a heterogeneity analysis of the hard treatment, based on the characteristics of analysts. Columns (1) and (2) interact algorithmic assignment with an indicator for new employees and find no statistically meaningful differential effect. The coefficient on the new-employee indicator is consistently negative, indicating that newer analysts complete fewer cases on time regardless of assignment method. Columns (3) and (4) interact the treatment with an indicator for analysts with five or more years of sector-specific experience. The interaction is negative but imprecise, providing at most suggestive evidence that more experienced analysts bear a larger productivity cost under algorithmic assignment. Columns (5) and (6) provide at most suggestive evidence of heterogeneity by case complexity: the interaction is significant without controls but becomes smaller and statistically insignificant once controls are added. Across all specifications, the coefficient on suggested dates remains positive and of similar magnitude to the main estimates.

Table F1: Heterogeneity Analysis of Productivity Results

	(1)	(2)	(3)	(4)	(5)	(6)
Algorithm Assignment	-0.043	-0.055	0.007	-0.060	0.008	-0.039
	(0.069)	(0.049)	(0.061)	(0.040)	(0.048)	(0.032)
New Employee	-0.168**	-0.040				
	(0.074)	(0.058)				
Experienced Worker			0.169***	0.005		
			(0.064)	(0.045)		
Complex Case					-0.015	-0.022
					(0.052)	(0.045)
Algorithm $\times$ New Employee	-0.007	-0.030				
	(0.088)	(0.059)				
Algorithm $\times$ Experienced			-0.110	-0.024		
			(0.083)	(0.055)		
Algorithm $\times$ Complex Case					-0.108*	-0.072
					(0.057)	(0.050)
Case with Suggested Date	0.039	0.047*	0.042	0.046*	0.038	0.047*
	(0.028)	(0.025)	(0.030)	(0.025)	(0.032)	(0.025)
Controls	No	Yes	No	Yes	No	Yes
Mean Outcome	0.404	0.404	0.404	0.404	0.404	0.404
N. Analysts	95	95	95	95	95	95
R2	0.212	0.293	0.196	0.293	0.192	0.294
N	1405	1405	1405	1405	1405	1405

Note: This table shows results of the experiment on the probability that a case was concluded before December 31, 2023, interacting the treatment variable with three characteristics. Columns (1)–(2) interact with new-employee status; Columns (3)–(4) with an indicator for analysts with five or more years of sector-specific experience; Columns (5)–(6) with case complexity. Odd-numbered Columns include only sector fixed effects (the randomization strata). Even-numbered Columns add case characteristics (month of assignment, complexity) and analyst-level characteristics (new employee dummy, master's degree dummy, high experience dummy, age-tercile dummies, analyst caseload, missing-covariates indicator). Standard errors are clustered at the analyst level. Significance levels: *10%, **5%, ***1%.

Table F2: Matching Mechanism: Full Coefficients for Complexity Assignment

	(1)	(2)	(3)
Experienced Worker	0.153***	0.030	0.139***
	(0.041)	(0.036)	(0.041)
New Employee	0.101*	-0.016	0.045
	(0.056)	(0.041)	(0.054)
Masters	-0.039	0.012	-0.095
	(0.034)	(0.035)	(0.058)
N. Cases	0.007***	0.008	0.005**
	(0.002)	(0.008)	(0.002)
Experienced W. x Algorithm			-0.114**
			(0.055)
New Employee x Algorithm			-0.056
			(0.070)
Masters x Algorithm			0.089
			(0.074)
N. Cases x Algorithm			0.016***
			(0.003)
Sample	Control	Algorithm	All
Mean Outcome	0.488	0.457	0.470
R2	0.281	0.192	0.222

Note: This table reports the full set of coefficients for the matching-quality exercise in Panel A of Table 6. The outcome is a binary indicator for whether a case is complex. Column (1) restricts the sample to control analysts, column (2) restricts the sample to algorithm-assigned analysts, and column (3) pools both groups and interacts analyst characteristics with algorithmic assignment. All columns include sector fixed effects and the controls described in the text. Standard errors are clustered at the analyst level. Significance levels: *10%, **5%, ***1%.